

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12Q 1/68		A2	(11) International Publication Number: WO 00/58519
			(43) International Publication Date: 5 October 2000 (05.10.00)
(21) International Application Number: PCT/US00/08440		(74) Agents: HOGLE, Doreen, M. et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).	
(22) International Filing Date: 30 March 2000 (30.03.00)			
(30) Priority Data: 60/127,248 31 March 1999 (31.03.99) US		(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(71) Applicants (for all designated States except US): WHITE-HEAD INSTITUTE FOR BIOMEDICAL RESEARCH [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US). AFFYMETRIX, INC. [US/US]; 3380 Central Expressway, Santa Clara, CA 95051 (US).		Published <i>Without international search report and to be republished upon receipt of that report.</i>	
(72) Inventors; and (75) Inventors/Applicants (for US only): ALTSHULER, David [US/US]; 69 Mason Terrace, Brookline, MA 02146 (US). CARGILL, Michele [US/US]; One Warwick Park #3, Cambridge, MA 02140 (US). DALEY, George, Q. [US/US]; 50 Young Road, Weston, MA 02493 (US). IRELAND, James, S. [US/US]; One Warwick Park, #3, Cambridge, MA 02140 (US). LANDER, Eric, S. [US/US]; 151 Bishop Allen Drive, Cambridge, MA 02139 (US). LIPSHUTZ, Robert, J. [US/US]; 970 Palo Alto Avenue, Palo Alto, CA 94301 (US). PATIL, Nila [US/US]; 780 West California Way, Woodside, CA 94602 (US). SKLAR, Pamela [US/US]; 68 Fuller Street, Brookline, MA 02446 (US).			
(54) Title: CHARATERIZATION OF SINGLE NUCLEOTIDE POLYMORPHISMS IN CODING REGIONS OF HUMAN GENES			
(57) Abstract			
<p>The invention provides nucleic acid segments of the human genome, particularly nucleic acid segments from the coding region of a gene, including polymorphic sites. Allele-specific primers and probes hybridizing to regions flanking or containing these sites are also provided. The nucleic acids, primers and probes are used in applications such as phenotype correlations, forensics, paternity testing, medicine and genetic analysis.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

CHARACTERIZATION OF SINGLE NUCLEOTIDE POLYMORPHISMS
IN CODING REGIONS OF HUMAN GENES

RELATED APPLICATION

This application claims the benefit of U.S. Provisional Application Serial
5 No. 60/127,248, filed March 31, 1999, the entire teachings of which are incorporated
herein by reference.

GOVERNMENT SUPPORT

The invention was supported, in whole or in part, by grant 5-P50-HG00098-
09 SNP from the National Institutes of Health (NCHGR) and grant 1-R01-
10 HL61774-01 from the National Institutes of Health (NHLBI). The Government has
certain rights in the invention.

BACKGROUND OF THE INVENTION

A major goal in human genetics is to understand the role of common genetic
variants in susceptibility to common diseases (N. Risch and K. Merikangas, *Science*,
15 273:1516 (1996.); *E. S. Lander, Science, 274:536 (1996); F.S. Collins, et al.,*
Science, 278:1580 (1997)). This requires assembling an extensive catalogue of
single-nucleotide polymorphisms (SNPs) and performing systematic association
studies for particular diseases.

The human population has relatively limited genetic diversity, reflecting its
20 young age and historically small size (F. J. Ayala *et. al., Proc. Natl. Acad. Sci.,*
91:6787 (1994)). Given the restricted nature of the allelic spectrum, some authors
have recently suggested that it should eventually be possible to collect all common
SNPs in the human population and have hypothesized that such common variants
may underlie much of the genetic risk of common disease (N. Risch and K.
25 Merikangas, *Science, 273:1516 (1996.); E. S. Lander, Science, 274:536 (1996); F.S.*
Collins, et al., Science, 278:1580 (1997)). This is in contrast to the situation for rare

Collins, *et al.*, *Science*, 278:1580 (1997)). This is in contrast to the situation for rare genetic diseases, which are primarily caused by a large number of distinct alleles that are recent, rare and highly penetrant. Important examples of associations to common (>1%) alleles include the ApoE4 allele in Alzheimer's disease, the Factor V^{Leiden} allele in deep-venous thrombosis, and the CCR5-Δ32 in resistance to HIV infection (A. M. Saunders *et al.*, *Neurology*, 43:1467 (1993); R. M. Bertina, *Nature*, 369:64 (1994); M. Dean *et al.*, *Science*, 273:1856 (1996)). The most relevant variants are likely to be those in coding and regulatory regions of genes.

SUMMARY OF THE INVENTION

As described herein, the nature of SNPs in the coding regions of human genes has been explored. SNPs were identified in 106 genes relevant to cardiovascular disease, endocrinology and neuropsychiatry, by screening an average of 114 independent alleles using two independent screening methods. To ensure high accuracy, all reported SNPs were confirmed by DNA sequencing. A total of 545 SNPs were identified, including 395 coding-regions SNPs (cSNPs) divided roughly equally between those causing synonymous and non-synonymous changes. The cSNPs most likely to influence disease, those that alter the amino acid sequence of the encoded protein, show strikingly different properties: they occur at a lower rate and with lower allele frequencies. This likely reflects selection acting against deleterious alleles during human evolution. The lower allele frequency of cSNPs has important implications for the number of chromosomes that must be sampled to construct a comprehensive catalogue of human cSNPs.

The invention relates to a gene which comprises a single nucleotide polymorphism at a specific location. In a particular embodiment the invention relates to the variant allele of a gene having a single nucleotide polymorphism, which variant allele differs from a reference allele by one nucleotide at the site(s) identified in Figures 5A-5Q. Complements of these nucleic acid segments are also included. The segments can be DNA or RNA, and can be double- or single-stranded. Segments can be, for example, 5-10, 5-15, 10-20, 5-25, 10-30, 10-50 or 10-100 bases long. The invention further relates to gene products encoded by genes and oligonucleotides of the invention.

The invention further provides allele-specific oligonucleotides that hybridize to a gene comprising a single nucleotide polymorphism or to the complement of the gene. These oligonucleotides can be probes or primers.

The invention further provides a method of analyzing a nucleic acid from an individual. The method determines which base is present at any one of the polymorphic sites shown in Figures 5A-5Q. Optionally, a set of bases occupying a set of the polymorphic sites shown in Figures 5A-5Q is determined. This type of analysis can be performed on a number of individuals, who are tested for the presence of a disease phenotype. The presence or absence of disease phenotype is then correlated with a base or set of bases present at the polymorphic site or sites in the individuals tested.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a graph showing minor allele frequency by polymorphism type. The percentage of cSNPs having minor allele frequency classified as low (<5%), medium (5-15%) or high (>15%) frequency is displayed for synonymous, non-synonymous and non-coding SNPs.

Figure 2 is a graph showing the distribution of nucleotide diversity. Normalized frequency of variant sites, θ , was calculated for the coding region of each gene. The graph shows the percentage of genes having θ in the indicated range.

Figures 3A and 3B are a table showing a summary of polymorphisms in 106 human genes described herein. Column 1 shows the name of the gene as used in Online Mendelian Inheritance in Man. Column 2 shows the number of coding base pairs screened. Column 3 shows the number of synonymous (or silent) polymorphisms identified. Column 4 shows the number of non-synonymous polymorphisms identified. Column 5 shows the number of non-coding base pairs screened. Column 6 shows the number of non-coding polymorphisms, including those in introns and untranslated regions (UTR), identified.

Figure 4 is a table showing polymorphism rates for different classes of sites.

Figures 5A-5Q are a table showing the specific polymorphisms identified in the genes studied as described herein. Column 1 shows the laboratory

designation for the polymorphism. Column 2 shows the name of the gene as used in Online Mendelian Inheritance in Man. Column 3 shows the reference nucleotide which occupies the polymorphic site in the reference allele. Column 4 shows the variant nucleotide which occupies the polymorphic site in the variant allele.

5 Column 5 shows the reference amino acid encoded by the codon which contains the polymorphic site in the reference allele. Column 6 shows the variant or alternate amino acid encoded by the codon which contains the polymorphic site in the variant allele. Column 7 indicates whether the polymorphism is located in the coding or non-coding region of the gene. Column 8 shows the assay number in which the

10 polymorphism was assessed. Columns 9 and 10 show the forward and reverse primers, respectively, which were used to identify the polymorphism. Column 11 shows the sequence of the gene used in the assay, with the polymorphic site indicated by brackets and the primers shown in capital letters. Column 12 shows the total number of nucleotides given in Column 11.

15 DETAILED DESCRIPTION OF THE INVENTION

There is a rich literature concerning nucleotide variation in model systems, particularly in *Drosophila* (E. N. Moriyama and J. R. Powell., *Mol. Biol. Evol.*, 13:261 (1996)), but sequence variation in human genes has been studied only in limited ways. A small number of studies have focused on individual genes (such as

20 beta-globin and lipoprotein lipase) in many individuals, and one study examined 49 genes by comparing two independent sequences deposited in public databases (R. M. Harding *et. al.*, *Am. J. Hum. Genet.*, 60:772 (1997); D. A. Nickerson *et. al.*, *Nature Genetics*, 19:233 (1998); W. H. Li and L.A. Sadler, *Genetics* 129:513 (1991)). To perform a more comprehensive survey, as described herein, a collection

25 of 106 genes were selected whose protein products play important roles in the cardiovascular, endocrine and neurological systems (Figures 3A-3B and Figures 5A-5Q). Gene sequences were obtained from the Genbank and TIGR databases. Where multiple sequence depositions were available, a consensus sequence was derived. Determination of coding sequence, untranslated regions and

30 intronic regions was based on annotation in the public database, although internal

checks were performed to ensure accurate determination of start and stop codons, open reading frames and the like.

The genes were chosen because of their relevance to common, clinically significant diseases, such as coronary artery disease, diabetes, and schizophrenia.

5 They encode proteins involved in coagulation, lipid metabolism, energy metabolism, neuroendocrine physiology, neurotransmission and central nervous system development. Variation in these genes was studied in a sample including Caucasians, African-Americans, African Pygmies and Asians, with an average of 114 chromosomes screened for each gene. Of the samples screened, 30 were from
10 Caucasian individuals, 14 from Asian, 10 African American and 7 Africans. The average number of individuals successfully screened for each gene was 57, with the precise number successfully screened varying among genes. Cell lines were obtained from Coriell Cell Repository, and DNA prepared according to standard protocols. In addition, 10 of the Caucasian samples used in this study were obtained
15 as anonymous blood samples from the Physician's Health Study (gift of Charles Hennekens and J. Michael Gaziano). The sample size provides greater than 65% power to detect alleles with frequency of 1%.

Overall, the sample of 114 chromosomes was screened for SNPs in a total of 195.4 kb, consisting of 135.8 kb of coding regions and 59.6 kb from adjacent
20 non-coding region (untranslated region (UTR) and introns). Sequences were amplified by the polymerase chain reaction (PCR) and screened by two independent methods. The first method involved hybridization of labeled PCR products to variant detector arrays (VDAs) (that is, high density DNA probe arrays containing oligonucleotides specific for the sequences under study) (M. Chee *et al.*, *Science*,
25 274:610 (1996); D. G. Wang *et al.*, *Science*, 280:1077 (1998)); variant sequences typically give rise to altered hybridization patterns. These chips contained variant detector arrays (VDA) (M. Chee *et al.*, *Science* 274:610 (1996)).

Using VDAs, candidate SNPs were identified using a combination of three algorithms followed by visual inspection. For each base position and strand queried
30 there are four VDA features: one contains the expected base (the reference sequence) in the central position and the other three features contain central substitution bases (in the background of the reference sequence). The base-calling

algorithm looked for positions at which hybridization to a substitution base gives a stronger signal than the reference base. The second algorithm (mutant fraction) examined the reference base and each one of the substitution bases in turn and calculates the fraction of signal present in the non-reference base. The final
5 algorithm (footprint detection) depends upon a loss of signal at the reference positions surrounding a nucleotide substitution. These algorithms are combined to yield a confidence score of "certain" or "likely" for each candidate polymorphism. Two analysts independently scored the data, and candidate polymorphisms found by either observer were included in subsequent confirmation tests. PCR assays
10 spanning each exon were designed using Primer 3.0 release 0.7. PCR was performed according to standard protocols, and assays destined to be hybridized to the same chip design were pooled together. Chip samples were prepared and hybridized as described in D.G. Wang *et al.* (*Science* 280:1077 (1998)), except that pools consisting of about 100 assays contained 5-6 µg of amplified material. In all,
15 854 assays (average size of 300 bp, covering 106 genes) were amplified from each individual and were hybridized to 12 distinct chip designs. The probe arrays were designed to query only the coding sequence for some genes, while other genes contained the entire mRNA and/or surrounding intron (Figures 3A-3B). The second method involved subjecting PCR products to Denaturing HPLC (dHPLC) (P. J.
20 Oefner and P. A. Underhill, *Am. J. Hum. Genet.*, 57:A266 (1995)) at a critical temperature; heterozygous individuals typically give rise to heteroduplex products with altered denaturation and migration properties.

Sequences were amplified as above except that the final extension in the PCR protocol was followed by denaturation and slow reannealing to allow
25 heteroduplex formation. A total of 6 µl of each individual PCR product was injected into Wave DNA Fragment Analysis System (Transgenomic). A total of 592 of the VDA assays (covering the 89 genes attempted with this method) were successfully screened by DHPLC. Only assays of >160 base pairs were used for DHPLC, because shorter assays performed unreliably for mutation detection. The
30 DHPLC parameters (percentage of acetonitrile, column temperature) used for each fragment were automatically calculated using a novel predictive algorithm, and DHPLC traces were analyzed using the clustering program ASH v2.0. A scoring

algorithm was developed based upon the similarity score by ASHv2.0 and contour of the elution profile.

Because both screening methods can generate to a significant number of false positives, it was important to confirm every reported SNP. Samples implicated
5 by either method as containing a candidate SNP were thus subjected to fluorescent dideoxy sequencing, either to confirm the presence of the SNP (in the case of the chip) or to identify and confirm the presence of the SNP (in the case of DHPLC). Such confirmation proved essential for eliminating false positives.

Candidate SNPs were either validated (if found by VDAs) or identified (if
10 implicated by DHPLC) by DNA sequencing. For this purpose, sequences were amplified with PCR primers tailed with standard M13 sequencing sites (-21 forward and -28 reverse) and conventional dye-primer sequencing was performed on ABI 377 sequencers. For candidate SNPs discovered by VDAs, one individual was chosen (a candidate homozygous variant, when available, or a candidate
15 heterozygote) and sequencing was performed on one strand to confirm by visual inspection the presence of the SNP at the indicated position. For amplicons found to be polymorphic by DHPLC, two individuals were selected representing each distinct elution pattern observed and were sequenced on both strands to discover the variant base or bases. Sequences were base-called by the Phred program, assembled
20 by the Phrap program, and polymorphism candidates were identified by the PolyPhred program (D. A. Nickerson *et. al.*, *NAR*, 25:2745 (1997). All results were visually inspected by at least two observers.

The overall false positive rate for VDAs was 45%. The rate was much lower (about 10%) for certain chip designs, synthesis protocols, and for candidate
25 polymorphisms scored as "certain." The false positive rate among fragments displaying an altered elution pattern by DHPLC was similar (40%). The false positive rates reflect the thresholds employed for declaring a candidate SNP, which were chosen to ensure high sensitivity.

A total of 545 SNPs were identified in the 195 kb surveyed, consisting of
30 150 non-coding SNPs and 395 cSNPs. Results from these studies are shown in the Figures. The complete data are available on the web site http://www.genome.wi.mit.edu/cvar_snps; access to this website can be gained

using the guestname "snp_pilot" and the password "noynek". In the future, access to this website may be available to the public, and thus, no guestname or password may be needed.

To directly determine the false-negative rate of the screen, conventional
 5 DNA sequencing was performed on ten of the genes (THPO, TBAX2R, PTHLH, IGF2, HTR2A, HTR1A, GHR, GABRB1, F10, and CYP11B1) spanning 25.2 kb in twenty individuals. Sequencing was performed on both strands using dye-primer chemistry and sequence traces were interpreted using PolyPhred (D.A. Nickerson *et al.*, *NAR*, 25:2745 (1997)). VDA analysis identified 85% of variants found by direct
 10 sequencing, while DHPLC identified 87% of the variants found by direct sequencing. In regions screened by both VDAs and DHPLC, the combination of the two methods identified 100% of the polymorphisms found by direct sequencing.

Overall, about one-third of individuals were screened with both methods, and one-third were screened with each of the two methods alone. (For some genes, the
 15 non-coding regions were screened only by DHPLC.) It is estimated that the false negative rate over the entire study to be about 15% for regions screened by one method, and negligible for sequences screened by both methods. The total number of true polymorphisms not identified is estimated to be less than 10%.

A SNP survey can be characterized in terms of either K, the observed
 20 number of variant sites, or p, the observed heterozygosity per bp. Because K increases with the number of chromosomes (n) studied and the total sequence length L, it is preferable to use the normalized number of variant sites

$$\hat{\theta} = K / \left(\sum_{i=1}^{n-1} i^{-1} \right) L \text{ which corrects for sample size. Under the neutral}$$

theory of molecular evolution and infinite sites model, θ and π are both estimators
 25 of the population genetic parameter $\theta = 4N\mu$ (Li, *Molecular Evolution*, Sinauer Associates (1997), Canada).

SNPs were found at a similar overall frequency in coding and non-coding regions. SNPs in coding region occurred at a frequency of 1 per 344 bp, corresponding to $\hat{\theta} = 5.47 \times 10^{-4}$ and $\pi = 5.07 \times 10^{-4}$. Interestingly, SNPs were
 30 observed in non-coding DNA at a similar frequency of 1 per 397 bp. The

normalized number of variant sites was $\theta = 4.93 \times 10^{-4}$, and the mean heterozygosity (π) = 5.05×10^{-4} (Figure 4). Calculations of π involve allele frequencies. Polymorphisms identified by DHPLC alone were excluded because we did not sequence all of the samples showing a variant DHPLC pattern and thus could not be certain of allele frequency. The estimates of π were thus based on 411 of 545 polymorphisms. Although the VDAs were designed for polymorphism discovery rather than genotyping, the estimated allele frequencies proved to be quite accurate. Specifically, genotyping assays (employing single-base extension assays) for 25 SNPs yielded allele frequencies that differed by an average of only 2% from those estimated on the basis of genotypes inferred from the VDA. For both classes, the similar values for θ and π is consistent with a population evolving according to neutral expectations.

The 395 cSNPs were roughly equally divided between synonymous (203 cSNPs) and non-synonymous (192 cSNPs) changes. Since approximately two-thirds of random mutations would alter an amino acid, the fact that non-synonymous cSNPs comprise slightly less than half of the cSNPs implies strong selection against amino-acid altering changes. To address this issue more directly, the nucleotide diversity was examined at four-fold degenerate sites, two-fold degenerate sites, and non-degenerate sites. Changes at four-fold degenerate sites produce only synonymous changes, while those at non-degenerate sites are always non-synonymous. Nucleotide diversity (θ) was 9.64×10^{-4} at four-fold degenerate sites, 6.85×10^{-4} at two-fold degenerate sites, and 3.70×10^{-4} at non-degenerate sites. Assuming that mutations occur at an equal rate at both classes of sites, non-synonymous variants survive to be detected in such a survey at only 38% of the rate of synonymous changes.

The force of selection is also evident in comparing non-synonymous cSNPs causing a non-conservative amino acid alteration with those causing a conservative amino-acid change. Conservative and non-conservative amino acid substitutions were defined for this analysis according to the BLOSUM62 matrix, used in sequence comparison (S. Henikoff and J. G. Henikoff, *PNAS*, 89:10915 (1992)). Conservative changes were those having a positive or neutral sign in the matrix, while non-conservative changes were those having a negative value. Non-conservative

- cSNPs represent only 36% of the non-synonymous cSNPs, whereas randomly distributed mutations would be expected to produce a higher proportion (52%) of non-conservative changes. The proportion of non-synonymous SNPs expected to cause a non-conservative amino acid substitution was determined based on the actual codon usage in the 106 genes studied, the known frequencies of transitions and transversions, and the definition of non-conservative changes employed in the BLOSUM62 matrix. This implies that non-conservative cSNPs survive to be detected in such a survey at only about half of the rate of conservative, non-synonymous cSNPs.
- 10 The various types of SNPs differ not only in the rate of their occurrence, but also in the frequency of their minor alleles. This can be seen in several ways. When SNPs are classified according to whether the frequency of the minor allele was high ($\geq 15\%$), intermediate (5-15%) or low ($\leq 5\%$), it is clear that the non-synonymous cSNPs were enriched in low frequency alleles compared to the rest of the collection
- 15 (Figure 1). The distribution of non-synonymous allele frequencies was significantly different than that of synonymous changes ($p=0.02$, Kolmogorov-Smirnov test). Indeed, more than half (58%) of non-synonymous cSNPs were found at a frequency below 5%, with this effect evident for both conservative and non-conservative substitutions.
- 20 The effect of selection can also be inferred by considering the average frequency of the minor allele: it is 8% for non-conservative cSNPs, 11% for conservative but non-synonymous cSNPs, and 14% for both synonymous cSNPs and non-coding SNPs. In addition, the lower allele frequency of non-synonymous cSNPs is reflected in the fact that the heterozygosity π is lower than the normalized
- 25 rate of variant sites θ for this class of SNPs (Figure 4). This divergence is in the direction predicted by the action of purifying selection, although it falls short of statistical significance. Tajima's D was non-significant. (F. Tajima, *Genetics*, 123:545 (1989).
- 30 The distribution of SNPs among the 106 genes was explored, with an eye toward detecting differential effects of selection among genes. The number of cSNPs per gene ranged from 37 for Factor V to 0 for thirteen of the genes, and the normalized rate, θ , similarly showed considerable variation (Figure 2). The

observed variation in nucleotide diversity is similar in magnitude to that observed for *Drosophila* (E. N. Moriyama and J. R. Powell., *Mol. Biol. Evol.*, 13:261 (1996)). Variation among genes could be due to many factors (D. J. Begun and C. F. Aquadro, *Nature*, 356:519 (1993); Nachman *et. al.*, *Genetics*, 150:1133 (1998)).

- 5 The fact that non-synonymous cSNPs show a somewhat wider variation than synonymous cSNPs (the coefficient of variation is 20% larger for the former class) is consistent with differences in selective constraints among loci, but the difference falls well below statistical significance. A variety of population genetic tests are available for testing selection at individual loci (M. L. Wayne and K. L. Simonson, 10 *Trends and Ecology and Evolution*, 13:236 (1998)).

The age of a SNP allele has important implications for its use in human genetic studies. Recently-occurring SNP alleles are more likely to show extensive linkage disequilibrium (retention of the ancestral haplotype on which they arose) as compared to older SNPs. Such linkage disequilibrium can provide a powerful tool 15 in identifying disease genes (E. S. Lander, N.J. Schork, *Science*, 265:2037 (1994)). Although the precise age of the SNPs could not be assessed from these studies, characterization of which allele preceded human speciation and which arose thereafter was sought. To determine the ancestral human allele, each corresponding gene was sequenced from the common chimpanzee (*P. troglodytes*). Each assay 20 used in the human survey was amplified from a single chimpanzee (DNA gift of Kristin Ardlie) and subjected to dye-primer sequencing on both strands. A single chimpanzee sample will accurately reveal the ancestral allele except in cases where the site has mutated and fixed during the chimpanzee evolution or is polymorphic in the chimpanzee population and happened to be homozygous for the non-ancestral 25 allele. These two cases are quite rare (probably less than 2%) and thus have been neglected for the purpose of estimating overall rates. A human allele was considered to be ancestral if it was present in the homozygous state in the chimpanzee sample. A total of 136 kb of chimpanzee sequence was obtained, revealing an inter-species divergence of 0.6% in the regions studied.

- 30 An elegant result in theoretical population genetics predicts that the probability that a neutral allele represents the ancestral state should be equal to its frequency in the population (G. A. Watterson and H. A. Guess, *Theoretical*

Population Biology, 11:141 (1977)). The minor allele should thus represent the ancestral state in a predictable proportion of cases. The ancestral allele and minor allele frequency was determined for 267 of the reported SNPs. For 3 of the 267 SNPs, the chimpanzee was homozygous for a third allele differing from both of the current human alleles. This is consistent with the overall 0.6% nucleotide sequence divergences seen between human and chimpanzee. Among polymorphisms with a minor allele frequency below 10%, the average allele frequency was 3% and the proportion that was ancestral was 7% (11/158) of cases. Among polymorphisms with minor alleles exceeding 10%, the mean frequency was 28% and the proportion that were ancestral was 32% (35/109). These results thus agree remarkably well with the theoretical prediction, providing the first reported test of this prediction in humans. It therefore follows that the minor SNP allele need not be the younger allele; this has implications for linkage disequilibrium mapping.

The distribution of SNPs among Caucasian, African-American, African and Asian samples was also examined. Although the vast majority of SNPs were seen in multiple groups, there was a statistically significant excess of SNPs that were seen in only one of the sub-groups. The probability that a SNP occurring $k > 1$ times in an overall sample of n individuals would be found entirely within a given subset of m individuals is $B(n,k)/B(m,k)$, where $B(x,y)$ is the binomial coefficient $x!/(x-y)!y!$. In this fashion, the probability that each individual SNP would be confined to a particular ethnic subgroup within the sample was calculated and these probabilities were summed to obtain the number of SNPs expected to be confined to the group within the sample. The fact that a SNP is found only within one group in the sample does not necessarily imply that it is private to that group within the general population, owing to the small sample size, but it can be used as an indication of substructure. The number of SNPs with $k > 1$ confined to the, African-Americans, African Pygmies, Caucasians, and Asians was 17, 17, 12, and 9, as compared to expectations of 3.02, 1.34, 8.62, and 1.81. Not surprisingly, the greatest excess was seen for SNPs found in the African-American and African samples. The presence of population substructure implies that construction of a comprehensive SNP database should employ a diverse set of DNA samples.

The results of this survey provide a fundamental description of sequence variation in the coding regions of human genes. These data indicate that two copies of a gene chosen from the human population will differ by roughly one base in 2 kb, corresponding to somewhat less than one heterozygous base within the coding region of a typical gene. In general, there are only a handful of such cSNPs per gene that exhibit allele frequencies of at least a few percent. Accounting for both the different rate and frequency of non-synonymous SNPs, only about 40% of these observed changes will alter the encoded amino acid. The action of purifying selection during human evolution is evident from the comparatively lower rate of non-synonymous cSNPs, and especially of those that create a non-conservative change. It is clear that non-synonymous cSNPs not only occur less often, but also have lower minor allele frequencies: 60% of non-synonymous cSNPs, the class likely to have the most dramatic effects on proteins, display a minor allele frequency below 5%.

The relative rarity of cSNPs has important implications for efforts to produce large catalogues of human variants. It has been proposed that most human SNPs could be found by performing shotgun sequencing on a handful of individuals (J. L. Weber and E. W. Myers, *Genome Research*, 7:401 (1997); J. C. Venter *et. al.*, *Science*, 280:1540 (1998)). Although such a project will surely identify many SNPs, results described herein suggest that the small sample size will likely fail to identify the vast majority of cSNPs likely to have the most important biological consequences, owing to their lower average allele frequencies. A comprehensive collection of the common, non-conservative cSNPs may require surveying 50-100 chromosomes. Because coding sequence represents only about 3% of the genome, it may prove inefficient to obtain such deep coverage of cSNPs by shotgun sequencing of genomic DNA. Instead, it may be more efficient to perform shotgun sequencing on cDNA libraries from multiple individuals or to amplify genes from multiple individuals, as done here.

Interestingly, a similar rate of polymorphism in coding and non-coding DNA was found. Furthermore, the observed rate of nucleotide diversity at four-fold degenerate sites was nearly twice that in adjacent non-coding regions, and over twice that at non-degenerate sites (Figure 4). Similar results have been reported for

Drosophila (E. N. Moriyama and J. R. Powell., *Mol. Biol. Evol.*, 13:261 (1996)) and for a smaller human data set by Li and Sadler (R. M. Harding *et. al.*, *Am. J. Hum. Genet.*, 60:772 (1997); D. A. Nickerson *et. al.*, *Nature Genetics*, 19:233 (1998); W. H. Li and L.A. Sadler, *Genetics* 129:513 (1991)), who observed over three times
5 the nucleotide diversity at four-fold degenerate sites ($\theta = 11 \times 10^{-4}$), as compared to that in both untranslated regions and non-degenerate sites ($\theta = 3 \times 10^{-4}$). These observations suggest that non-coding DNA adjacent to coding regions may be functionally constrained to a surprising degree.

SNPs can be used to search for genes underlying complex traits in two
10 distinct ways: linkage disequilibrium (LD) studies and association studies (E. S. Lander, N.J. Schork, *Science*, 265:2037 (1994)). Genome-wide LD studies involve using a dense collection of SNPs as markers to search for an ancestral haplotype carrying a disease-susceptibility allele. Such studies cannot be undertaken without the availability of an extremely dense SNP map and their potential for success
15 depends sensitively on many population genetic assumptions. Association studies are more straightforward because they directly test the hypothesis that a specific SNP increases disease risk. They make few assumptions, and require only the availability of a suitable database of appropriate SNPs. In the near term, focusing on cSNPs is likely to be most productive inasmuch as the class is easily recognized (in
20 contrast to regulatory polymorphisms) and is likely to contain a significant proportion of the disease-susceptibility alleles.

The present invention relates to a gene which comprises a single nucleotide polymorphism (SNP) at a specific location. The gene which includes the SNP has at least two alleles, referred to herein as the reference allele and the variant allele. The
25 reference allele (prototypical or wild type allele) has been designated arbitrarily and typically corresponds to the nucleotide sequence of the gene which has been deposited with GenBank or TIGR under a given Accession number. The variant allele differs from the reference allele by one at least one nucleotide at the site(s) identified in Figures 5A-5Q. The present invention also relates to variant
30 alleles of the described genes and to complements of the variant alleles. The invention further relates to portions of the variant alleles and portions of complements of the variant alleles which comprise (encompass) the site of the SNP

and are at least 5 nucleotides in length. Portions can be, for example, 5-10, 5-15, 10-20, 5-25, 10-30, 10-50 or 10-100 bases long. For example, a portion of a variant allele which is 5 nucleotides in length includes the single nucleotide polymorphism (the nucleotide which differs from the reference allele at that site) and four

5 additional nucleotides which flank the site in the variant allele. These nucleotides can be on one or both sides of the polymorphism. Polymorphisms which are the subject of this invention are defined in Figures 5A-5QQQQQQQ with respect to the reference sequence deposited in GenBank under the Accession number indicated. For example, the invention relates to a portion of a gene (e.g., AADC) having a
10 partial nucleotide sequence as shown in Figures 5A-5QQQQQQQ comprising a single nucleotide polymorphism at a specific position. The reference nucleotide for AADC is shown in column 3 and the variant nucleotide is shown in column 4 of Figures 5A-5QQQQQQQ. The nucleotide sequences of the invention can be double- or single-stranded.

15 The invention further provides allele-specific oligonucleotides that hybridize to a gene comprising a single nucleotide polymorphism or to the complement of the gene. These oligonucleotides can be probes or primers.

The invention further provides a method of analyzing a nucleic acid from an individual. The method determines which base is present at any one of the
20 polymorphic sites shown in Figures 5A-5QQQQQQQ. Optionally, a set of bases occupying a set of the polymorphic sites shown in Figures 5A-5QQQQQQQ is determined. This type of analysis can be performed on a number of individuals, who are tested for the presence of a disease phenotype. The presence or absence of disease phenotype is then correlated with a base or set of bases present at the
25 polymorphic site or sites in the individuals tested.

An oligonucleotide of this invention can be DNA or RNA, and single- or double-stranded. Oligonucleotides can be naturally occurring or synthetic, but are typically prepared by synthetic means. Preferred oligonucleotides of the invention include segments of DNA, or their complements, which include any one of the
30 polymorphic sites shown in Figures 5A-5QQQQQQQ. The segments can be between 5 and 250 bases, and, in specific embodiments, are between 5-10, 5-20, 10-20, 10-50, 20-50 or 10-100 bases. The polymorphic site can occur within any

position of the segment. The segments can be from any of the allelic forms of DNA shown in Figures 5A-5Q.

As used herein, the terms "nucleotide" and "nucleic acid" are intended to be equivalent. The terms "nucleotide sequence", "nucleic acid sequence", "nucleic acid molecule" and "segment" are intended to be equivalent.

Hybridization probes are oligonucleotides which bind in a base-specific manner to a complementary strand of nucleic acid. Such probes include peptide nucleic acids, as described in Nielsen *et al.*, *Science* 254, 1497-1500 (1991). Probes can be any length suitable for specific hybridization to the target nucleic acid sequence. The most appropriate length of the probe may vary depending upon the hybridization method in which it is being used; for example, particular lengths may be more appropriate for use in microfabricated arrays, while other lengths may be more suitable for use in classical hybridization methods. Suitable probes and primers can range from about 5 nucleotides to about 30 nucleotides in length. For example, probes and primers can be 5, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 25, 26, 28 or 30 nucleotides in length. The probe or primer preferably contains at least one polymorphic site occupied by any of the possible variant nucleotides. The nucleotide sequence can correspond to the coding sequence of the allele or to the complement of the coding sequence of the allele.

As used herein, the term "primer" refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis under appropriate conditions (*e.g.*, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer, but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template, but must be sufficiently complementary to hybridize with a template. The term primer site refers to the area of the target DNA to which a primer hybridizes. The term primer pair refers to a set of primers including a 5' (upstream) primer that hybridizes with the 5'

end of the DNA sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

As used herein, linkage describes the tendency of genes, alleles, loci or genetic markers to be inherited together as a result of their location on the same
5 chromosome. It can be measured by percent recombination between the two genes, alleles, loci or genetic markers.

As used herein, polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. A polymorphic marker or site is the locus at which divergence occurs. Preferred
10 markers have at least two alleles, each occurring at frequency of greater than 1%, and more preferably greater than 10% or 20% of a selected population. A polymorphic locus may be as small as one base pair. Polymorphic markers include restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide
15 repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements such as Alu. The first identified allelic form is arbitrarily designated as the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wildtype form. Diploid organisms may be homozygous or heterozygous
20 for allelic forms. A diallelic or biallelic polymorphism has two forms. A triallelic polymorphism has three forms.

By altering amino acid sequence, SNPs may alter the function of the encoded proteins. The discovery of the SNP facilitates biochemical analysis of the variants and the development of assays to characterize the variants and to screen for
25 pharmaceutical that would interact directly with on or another form of the protein. SNPs (including silent SNPs) may also alter the regulation of the gene at the transcriptional or post-transcriptional level. SNPs (including silent SNPs) also enable the development of specific DNA, RNA, or protein-based diagnostics that detect the presence or absence of the polymorphism in particular conditions.

30 A single nucleotide polymorphism occurs at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site

is usually preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than 1/100 or 1/1000 members of the populations).

A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa. Single nucleotide polymorphisms can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele. Typically the polymorphic site is occupied by a base other than the reference base. For example, where the reference allele contains the base "T" at the polymorphic site, the altered allele can contain a "C", "G" or "A" at the polymorphic site.

Hybridizations are usually performed under stringent conditions, for example, at a salt concentration of no more than 1 M and a temperature of at least 25°C. For example, conditions of 5X SSPE (750 mM NaCl, 50 mM NaPhosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30°C, or equivalent conditions, are suitable for allele-specific probe hybridizations. Equivalent conditions can be determined by varying one or more of the parameters given as an example, as known in the art, while maintaining a similar degree of identity or similarity between the target nucleotide sequence and the primer or probe used.

The term "isolated" is used herein to indicate that the material in question exists in a physical milieu distinct from that in which it occurs in nature. For example, an isolated nucleic acid of the invention may be substantially isolated with respect to the complex cellular milieu in which it naturally occurs. In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstance, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 percent (on a molar basis) of all macromolecular species present.

I. Analysis of Polymorphisms

A. Preparation of Samples

Polymorphisms are detected in a target nucleic acid from an individual being analyzed. For assay of genomic DNA, virtually any biological sample (other than
5 pure red blood cells) is suitable. For example, convenient tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. For assay of cDNA or mRNA, the tissue sample must be obtained from an organ in which the target nucleic acid is expressed. For example, if the target nucleic acid is a cytochrome P450, the liver is a suitable source.

10 Many of the methods described below require amplification of DNA from target samples. This can be accomplished by e.g., PCR. *See generally PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, *et al.*, Academic Press, San Diego, CA, 1990); Mattila *et al.*, *Nucleic Acids Res.* 19, 4967 (1991); Eckert *et al.*, *PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson *et al.*, IRL Press, Oxford); and U.S. Patent 4,683,202.

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4, 560 (1989), Landegren *et al.*, *Science* 241, 1077
20 (1988), transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86, 1173 (1989)), and self-sustained sequence replication (Guatelli *et al.*, *Proc. Nat. Acad. Sci. USA*, 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and
25 double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

B. Detection of Polymorphisms in Target DNA

There are two distinct types of analysis of target DNA for detecting polymorphisms. The first type of analysis, sometimes referred to as de novo
30 characterization, is carried out to identify polymorphic sites not previously characterized (i.e., to identify new polymorphisms). This analysis compares target

sequences in different individuals to identify points of variation, i.e., polymorphic sites. By analyzing groups of individuals representing the greatest ethnic diversity among humans and greatest breed and species variety in plants and animals, patterns characteristic of the most common alleles/haplotypes of the locus can be identified, and the frequencies of such alleles/haplotypes in the population can be determined. Additional allelic frequencies can be determined for subpopulations characterized by criteria such as geography, race, or gender. The de novo identification of polymorphisms of the invention is described in the Examples section. The second type of analysis determines which form(s) of a characterized (known) polymorphism are present in individuals under test. There are a variety of suitable procedures, which are discussed in turn.

1. Allele-Specific Probes

The design and use of allele-specific probes for analyzing polymorphisms is described by e.g., Saiki *et al.*, *Nature* 324, 163-166 (1986); Dattagupta, EP 235,726, Saiki, WO 89/11548. Allele-specific probes can be designed that hybridize to a segment of target DNA from one individual but do not hybridize to the corresponding segment from another individual due to the presence of different polymorphic forms in the respective segments from the two individuals. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Some probes are designed to hybridize to a segment of target DNA such that the polymorphic site aligns with a central position (e.g., in a 15-mer at the 7 position; in a 16-mer, at either the 8 or 9 position) of the probe. This design of probe achieves good discrimination in hybridization between different allelic forms.

Allele-specific probes are often used in pairs, one member of a pair showing a perfect match to a reference form of a target sequence and the other member showing a perfect match to a variant form. Several pairs of probes can then be immobilized on the same support for simultaneous analysis of multiple polymorphisms within the same target sequence.

2. Tiling Arrays

The polymorphisms can also be identified by hybridization to nucleic acid arrays, some examples of which are described in WO 95/11995. One form of such arrays is described in the Examples section in connection with de novo identification
5 of polymorphisms. The same array or a different array can be used for analysis of characterized polymorphisms. WO 95/11995 also describes subarrays that are optimized for detection of a variant form of a precharacterized polymorphism. Such a subarray contains probes designed to be complementary to a second reference sequence, which is an allelic variant of the first reference sequence. The second
10 group of probes is designed by the same principles as described in the Examples, except that the probes exhibit complementarity to the second reference sequence. The inclusion of a second group (or further groups) can be particularly useful for analyzing short subsequences of the primary reference sequence in which multiple mutations are expected to occur within a short distance commensurate with the
15 length of the probes (e.g., two or more mutations within 9 to 21 bases).

3. Allele-Specific Primers

An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of an allelic form to which the primer exhibits perfect complementarity. See Gibbs, *Nucleic Acid Res.* 17, 2427-2448
20 (1989). This primer is used in conjunction with a second primer which hybridizes at a distal site. Amplification proceeds from the two primers, resulting in a detectable product which indicates the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect
25 complementarity to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing to elongation from the primer (see, e.g., WO 93/22456).

4. Direct-Sequencing

The direct analysis of the sequence of polymorphisms of the present invention can be accomplished using either the dideoxy chain termination method or the Maxam Gilbert method (see Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989); Zyskind *et al.*, *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988)).

5. Denaturing Gradient Gel Electrophoresis

Amplification products generated using the polymerase chain reaction can be analyzed by the use of denaturing gradient gel electrophoresis. Different alleles can be identified based on the different sequence-dependent melting properties and electrophoretic migration of DNA in solution. Erlich, ed., *PCR Technology, Principles and Applications for DNA Amplification*, (W.H. Freeman and Co, New York, 1992), Chapter 7.

6. Single-Strand Conformation Polymorphism Analysis

Alleles of target sequences can be differentiated using single-strand conformation polymorphism analysis, which identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita *et al.*, *Proc. Nat. Acad. Sci.* 86, 2766-2770 (1989). Amplified PCR products can be generated as described above, and heated or otherwise denatured, to form single stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures which are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products can be related to base-sequence differences between alleles of target sequences.

II. Methods of Use

After determining polymorphic form(s) present in an individual at one or more polymorphic sites, this information can be used in a number of methods.

A. Forensics

Determination of which polymorphic forms occupy a set of polymorphic sites in an individual identifies a set of polymorphic forms that distinguishes the individual. See generally National Research Council, *The Evaluation of Forensic*
5 *DNA Evidence* (Eds. Pollard *et al.*, National Academy Press, DC, 1996). The more sites that are analyzed, the lower the probability that the set of polymorphic forms in one individual is the same as that in an unrelated individual. Preferably, if multiple sites are analyzed, the sites are unlinked. Thus, polymorphisms of the invention are often used in conjunction with polymorphisms in distal genes. Preferred
10 polymorphisms for use in forensics are biallelic because the population frequencies of two polymorphic forms can usually be determined with greater accuracy than those of multiple polymorphic forms at multi-allelic loci.

The capacity to identify a distinguishing or unique set of forensic markers in an individual is useful for forensic analysis. For example, one can determine
15 whether a blood sample from a suspect matches a blood or other tissue sample from a crime scene by determining whether the set of polymorphic forms occupying selected polymorphic sites is the same in the suspect and the sample. If the set of polymorphic markers does not match between a suspect and a sample, it can be concluded (barring experimental error) that the suspect was not the source of the
20 sample. If the set of markers does match, one can conclude that the DNA from the suspect is consistent with that found at the crime scene. If frequencies of the polymorphic forms at the loci tested have been determined (e.g., by analysis of a suitable population of individuals), one can perform a statistical analysis to determine the probability that a match of suspect and crime scene sample would
25 occur by chance.

$p(\text{ID})$ is the probability that two random individuals have the same polymorphic or allelic form at a given polymorphic site. In biallelic loci, four genotypes are possible: AA, AB, BA, and BB. If alleles A and B occur in a haploid genome of the organism with frequencies x and y , the probability of each genotype
30 in a diploid organism is (see WO 95/12607):

$$\begin{aligned}\text{Homozygote: } p(\text{AA}) &= x^2 \\ \text{Homozygote: } p(\text{BB}) &= y^2 = (1-x)^2\end{aligned}$$

Single Heterozygote: $p(AB)=p(BA)=xy = x(1-x)$

Both Heterozygotes: $p(AB+BA)=2xy = 2x(1-x)$

The probability of identity at one locus (i.e, the probability that two individuals, picked at random from a population will have identical polymorphic forms at a given locus) is given by the equation:

$$p(ID) = (x^2)^2 + (2xy)^2 + (y^2)^2.$$

These calculations can be extended for any number of polymorphic forms at a given locus. For example, the probability of identity $p(ID)$ for a 3-allele system where the alleles have the frequencies in the population of x , y and z , respectively, is equal to the sum of the squares of the genotype frequencies:

$$p(ID) = x^4 + (2xy)^2 + (2yz)^2 + (2xz)^2 + z^4 + y^4$$

In a locus of n alleles, the appropriate binomial expansion is used to calculate $p(ID)$ and $p(exc)$.

The cumulative probability of identity (cum $p(ID)$) for each of multiple unlinked loci is determined by multiplying the probabilities provided by each locus.

$$\text{cum } p(ID) = p(ID1)p(ID2)p(ID3).... p(IDn)$$

The cumulative probability of non-identity for n loci (i.e. the probability that two random individuals will be different at 1 or more loci) is given by the equation:

$$\text{cum } p(\text{nonID}) = 1 - \text{cum } p(ID).$$

If several polymorphic loci are tested, the cumulative probability of non-identity for random individuals becomes very high (e.g., one billion to one). Such probabilities can be taken into account together with other evidence in determining the guilt or innocence of the suspect.

B. Paternity Testing

The object of paternity testing is usually to determine whether a male is the father of a child. In most cases, the mother of the child is known and thus, the mother's contribution to the child's genotype can be traced. Paternity testing investigates whether the part of the child's genotype not attributable to the mother is consistent with that of the putative father. Paternity testing can be performed by analyzing sets of polymorphisms in the putative father and the child.

If the set of polymorphisms in the child attributable to the father does not match the set of polymorphisms of the putative father, it can be concluded, barring experimental error, that the putative father is not the real father. If the set of polymorphisms in the child attributable to the father does match the set of

5 polymorphisms of the putative father, a statistical calculation can be performed to determine the probability of coincidental match.

The probability of parentage exclusion (representing the probability that a random male will have a polymorphic form at a given polymorphic site that makes him incompatible as the father) is given by the equation (see WO 95/12607):

10
$$p(\text{exc}) = xy(1-xy)$$

where x and y are the population frequencies of alleles A and B of a biallelic polymorphic site.

(At a triallelic site $p(\text{exc}) = xy(1-xy) + yz(1-yz) + xz(1-xz) + 3xyz(1-xyz)$), where x, y and z are the respective population frequencies of alleles A, B and C).

15 The probability of non-exclusion is

$$p(\text{non-exc}) = 1 - p(\text{exc})$$

The cumulative probability of non-exclusion (representing the value obtained when n loci are used) is thus:

$$\text{cum } p(\text{non-exc}) = p(\text{non-exc1})p(\text{non-exc2})p(\text{non-exc3})\dots p(\text{non-excn})$$

20 The cumulative probability of exclusion for n loci (representing the probability that a random male will be excluded)

$$\text{cum } p(\text{exc}) = 1 - \text{cum } p(\text{non-exc}).$$

If several polymorphic loci are included in the analysis, the cumulative probability of exclusion of a random male is very high. This probability can be

25 taken into account in assessing the liability of a putative father whose polymorphic marker set matches the child's polymorphic marker set attributable to his/her father.

C. Correlation of Polymorphisms with Phenotypic Traits

The polymorphisms of the invention may contribute to the phenotype of an organism in different ways. Some polymorphisms occur within a protein coding

30 sequence and contribute to phenotype by affecting protein structure. The effect may be neutral, beneficial or detrimental, or both beneficial and detrimental, depending

on the circumstances. For example, a heterozygous sickle cell mutation confers resistance to malaria, but a homozygous sickle cell mutation is usually lethal. Other polymorphisms occur in noncoding regions but may exert phenotypic effects indirectly via influence on replication, transcription, and translation. A single
5 polymorphism may affect more than one phenotypic trait. Likewise, a single phenotypic trait may be affected by polymorphisms in different genes. Further, some polymorphisms predispose an individual to a distinct mutation that is causally related to a certain phenotype.

Phenotypic traits include diseases that have known but hitherto unmapped
10 genetic components. Phenotypic traits also include symptoms of, or susceptibility to, multifactorial diseases of which a component is or may be genetic, such as autoimmune diseases, inflammation, cancer, diseases of the nervous system, and infection by pathogenic microorganisms. Some examples of diseases which can be treated or diagnosed as described herein include, but are not limited to,
15 bradyarrhythmias, tachyarrhythmias, heart failure, such as congestive heart failure, congenital heart disease, rheumatic fever, valvular heart disease, cardiomyopathies, myocarditides, pericardial diseases, cardiac tumors, cardiac manifestations of systemic diseases, and traumatic cardiac injury. Other disorders include atherosclerosis, acute myocardial infarction, ischemic heart disease, hypertensive
20 vascular disease, disorders of the aorta, vascular diseases of the extremities, vessel wall disorders, such as various forms of thrombocytopenia, von Willebrand's disease and drug-induced platelet dysfunction, and homeostatic disorders relating to vessel disease and associated bleeding. Also suitable are thrombotic thrombocytopenic purpura, hemolytic-uremic syndrome, Henoch-Schönlein purpura, capillary fragility,
25 vascular purpura, metabolic and inflammatory disorders, such as those induced by rickettsiae and certain drugs, such as sulfonamides, aortic aneurysm, aortic dissection, aortic occlusion, aortitis, atherosclerosis, coronary artery disease, angina, myocardial infarction, thrombosis, hemostatic and coagulation disorders, hypertension and hypotension. Other disorders include transplant accelerated
30 vascular restenosis following balloon angioplasty, Raynaud's disease and acrocyanosis.

Additional disorders include, but are not limited to, disorders of neurodegeneration characterized by astrocyte hypertrophy including gliosis, Pick's disease, aceroplasminemia, portal-systemic encephalopathy, frontal lobe dementia and inherited and acquired ataxias, neurodegenerative diseases of other etiology including progressive supranuclear palsy, primary progressive aphasia, cortical basal degeneration, Alzheimer's disease, Huntington's disease, and Parkinson's disease, retinitis pigmentosa and amyotrophic lateral sclerosis. Other disorders include epilepsy, stroke, defects of neural migration and differentiation, including Miller-Dieker lissencephaly syndrome, and cancer of the brain including astrocytomas and gliomas, as well as psychological disorders such as schizophrenia.

Phenotypic traits also include characteristics such as longevity, appearance (e.g., baldness, obesity), strength, speed, endurance, fertility, and susceptibility or receptivity to particular drugs or therapeutic treatments.

The correlation of one or more polymorphisms with phenotypic traits can be facilitated by knowledge of the gene product of the wild type (reference) gene. The genes in which cSNPs of the present invention have been identified are genes which have been previously sequenced and characterized in one of their allelic forms.

Correlation is performed for a population of individuals who have been tested for the presence or absence of a phenotypic trait of interest and for polymorphic markers sets. To perform such analysis, the presence or absence of a set of polymorphisms (i.e. a polymorphic set) is determined for a set of the individuals, some of whom exhibit a particular trait, and some of which exhibit lack of the trait. The alleles of each polymorphism of the set are then reviewed to determine whether the presence or absence of a particular allele is associated with the trait of interest. Correlation can be performed by standard statistical methods such as a K-squared test and statistically significant correlations between polymorphic form(s) and phenotypic characteristics are noted. For example, it might be found that the presence of allele A1 at polymorphism A correlates with heart disease. As a further example, it might be found that the combined presence of allele A1 at polymorphism A and allele B1 at polymorphism B correlates with increased milk production of a farm animal.

Such correlations can be exploited in several ways. In the case of a strong correlation between a set of one or more polymorphic forms and a disease for which treatment is available, detection of the polymorphic form set in a human or animal patient may justify immediate administration of treatment, or at least the institution of regular monitoring of the patient. Detection of a polymorphic form correlated with serious disease in a couple contemplating a family may also be valuable to the couple in their reproductive decisions. For example, the female partner might elect to undergo in vitro fertilization to avoid the possibility of transmitting such a polymorphism from her husband to her offspring. In the case of a weaker, but still statistically significant correlation between a polymorphic set and human disease, immediate therapeutic intervention or monitoring may not be justified. Nevertheless, the patient can be motivated to begin simple life-style changes (e.g., diet, exercise) that can be accomplished at little cost to the patient but confer potential benefits in reducing the risk of conditions to which the patient may have increased susceptibility by virtue of variant alleles. Identification of a polymorphic set in a patient correlated with enhanced receptiveness to one of several treatment regimes for a disease indicates that this treatment regime should be followed.

For animals and plants, correlations between characteristics and phenotype are useful for breeding for desired characteristics. For example, Beitz *et al.*, US 5,292,639 discuss use of bovine mitochondrial polymorphisms in a breeding program to improve milk production in cows. To evaluate the effect of mtDNA D-loop sequence polymorphism on milk production, each cow was assigned a value of 1 if variant or 0 if wildtype with respect to a prototypical mitochondrial DNA sequence at each of 17 locations considered. Each production trait was analyzed individually with the following animal model:

$$Y_{ijkpn} = \mu + YS_i + P_j + X_k + \beta_1 + \dots \beta_{17} + PE_n + a_n + e_p$$

where Y_{ijkpn} is the milk, fat, fat percentage, SNF, SNF percentage, energy concentration, or lactation energy record; μ is an overall mean; YS_i is the effect common to all cows calving in year-season; X_k is the effect common to cows in either the high or average selection line; β_1 to β_{17} are the binomial regressions of production record on mtDNA D-loop sequence polymorphisms; PE_n is permanent environmental effect common to all records of cow n ; a_n is effect of animal n and is

composed of the additive genetic contribution of sire and dam breeding values and a Mendelian sampling effect; and e_p is a random residual. It was found that eleven of seventeen polymorphisms tested influenced at least one production trait. Bovines having the best polymorphic forms for milk production at these eleven loci are used
5 as parents for breeding the next generation of the herd.

D. Genetic Mapping of Phenotypic Traits

The previous section concerns identifying correlations between phenotypic traits and polymorphisms that directly or indirectly contribute to those traits. The present section describes identification of a physical linkage between a genetic locus
10 associated with a trait of interest and polymorphic markers that are not associated with the trait, but are in physical proximity with the genetic locus responsible for the trait and co-segregate with it. Such analysis is useful for mapping a genetic locus associated with a phenotypic trait to a chromosomal position, and thereby cloning gene(s) responsible for the trait. See Lander *et al.*, *Proc. Natl. Acad. Sci. (USA)* 83,
15 7353-7357 (1986); Lander *et al.*, *Proc. Natl. Acad. Sci. (USA)* 84, 2363-2367 (1987); Donis-Keller *et al.*, *Cell* 51, 319-337 (1987); Lander *et al.*, *Genetics* 121, 185-199 (1989)). Genes localized by linkage can be cloned by a process known as directional cloning. See Wainwright, *Med. J. Australia* 159, 170-174 (1993); Collins, *Nature Genetics* 1, 3-6 (1992).

20 Linkage studies are typically performed on members of a family. Available members of the family are characterized for the presence or absence of a phenotypic trait and for a set of polymorphic markers. The distribution of polymorphic markers in an informative meiosis is then analyzed to determine which polymorphic markers co-segregate with a phenotypic trait. See, e.g., Kerem *et al.*, *Science* 245, 1073-1080
25 (1989); Monaco *et al.*, *Nature* 316, 842 (1985); Yamoka *et al.*, *Neurology* 40, 222-226 (1990); Rossiter *et al.*, *FASEB Journal* 5, 21-27 (1991).

Linkage is analyzed by calculation of LOD (log of the odds) values. A lod value is the relative likelihood of obtaining observed segregation data for a marker and a genetic locus when the two are located at a recombination fraction θ , versus
30 the situation in which the two are not linked, and thus segregating independently (Thompson & Thompson, *Genetics in Medicine* (5th ed, W.B. Saunders Company,

Philadelphia, 1991); Strachan, "Mapping the human genome" in *The Human Genome* (BIOS Scientific Publishers Ltd, Oxford), Chapter 4). A series of likelihood ratios are calculated at various recombination fractions (θ), ranging from $\theta = 0.0$ (coincident loci) to $\theta = 0.50$ (unlinked). Thus, the likelihood at a given value of θ is: probability of data if loci linked at θ to probability of data if loci 5 unlinked. The computed likelihoods are usually expressed as the \log_{10} of this ratio (i.e., a lod score). For example, a lod score of 3 indicates 1000:1 odds against an apparent observed linkage being a coincidence. The use of logarithms allows data collected from different families to be combined by simple addition. Computer 10 programs are available for the calculation of lod scores for differing values of θ (e.g., LIPED, MLINK (Lathrop, *Proc. Nat. Acad. Sci. (USA)* 81, 3443-3446 (1984)). For any particular lod score, a recombination fraction may be determined from mathematical tables. See Smith *et al.*, *Mathematical tables for research workers in human genetics* (Churchill, London, 1961); Smith, *Ann. Hum. Genet.* 32, 127-150 15 (1968). The value of θ at which the lod score is the highest is considered to be the best estimate of the recombination fraction.

Positive lod score values suggest that the two loci are linked, whereas negative values suggest that linkage is less likely (at that value of θ) than the possibility that the two loci are unlinked. By convention, a combined lod score of 20 +3 or greater (equivalent to greater than 1000:1 odds in favor of linkage) is considered definitive evidence that two loci are linked. Similarly, by convention, a negative lod score of -2 or less is taken as definitive evidence against linkage of the two loci being compared. Negative linkage data are useful in excluding a chromosome or a segment thereof from consideration. The search focuses on the 25 remaining non-excluded chromosomal locations.

III. Modified Polypeptides and Gene Sequences

The invention further provides variant forms of nucleic acids and corresponding proteins. The nucleic acids comprise one of the sequences described in Figures 5A-5Q, column 11, in which the polymorphic position is 30 occupied by one of the alternative bases for that position. Some nucleic acids encode full-length variant forms of proteins. Similarly, variant proteins have the

prototypical amino acid sequences encoded by nucleic acid sequences shown in Figures 5A-5Q, column 11, (read so as to be in-frame with the full-length coding sequence of which it is a component) except at an amino acid encoded by a codon including one of the polymorphic positions shown in Figures 5A-

- 5 5Q. That position is occupied by the amino acid coded by the corresponding codon in any of the alternative forms shown in Figures 5A-5Q.

Variant genes can be expressed in an expression vector in which a variant gene is operably linked to a native or other promoter. Usually, the promoter is a eukaryotic promoter for expression in a mammalian cell. The transcription regulation sequences typically include a heterologous promoter and optionally an enhancer which is recognized by the host. The selection of an appropriate promoter, for example trp, lac, phage promoters, glycolytic enzyme promoters and tRNA promoters, depends on the host selected. Commercially available expression vectors can be used. Vectors can include host-recognized replication systems, amplifiable genes, selectable markers, host sequences useful for insertion into the host genome, and the like.

The means of introducing the expression construct into a host cell varies depending upon the particular construction and the target host. Suitable means include fusion, conjugation, transfection, transduction, electroporation or injection, as described in Sambrook, *supra*. A wide variety of host cells can be employed for expression of the variant gene, both prokaryotic and eukaryotic. Suitable host cells include bacteria such as *E. coli*, yeast, filamentous fungi, insect cells, mammalian cells, typically immortalized, *e.g.*, mouse, CHO, human and monkey cell lines and derivatives thereof. Preferred host cells are able to process the variant gene product to produce an appropriate mature polypeptide. Processing includes glycosylation, ubiquitination, disulfide bond formation, general post-translational modification, and the like.

The protein may be isolated by conventional means of protein biochemistry and purification to obtain a substantially pure product, *i.e.*, 80, 95 or 99% free of cell component contaminants, as described in Jacoby, *Methods in Enzymology* Volume 104, Academic Press, New York (1984); Scopes, *Protein Purification, Principles*

and Practice, 2nd Edition, Springer-Verlag, New York (1987); and Deutscher (ed), *Guide to Protein Purification, Methods in Enzymology*, Vol. 182 (1990). If the protein is secreted, it can be isolated from the supernatant in which the host cell is grown. If not secreted, the protein can be isolated from a lysate of the host cells.

- 5 The invention further provides transgenic nonhuman animals capable of expressing an exogenous variant gene and/or having one or both alleles of an endogenous variant gene inactivated. Expression of an exogenous variant gene is usually achieved by operably linking the gene to a promoter and optionally an enhancer, and microinjecting the construct into a zygote. *See Hogan et al.*,
10 "Manipulating the Mouse Embryo, A Laboratory Manual," Cold Spring Harbor Laboratory. Inactivation of endogenous variant genes can be achieved by forming a transgene in which a cloned variant gene is inactivated by insertion of a positive selection marker. *See Capecchi, Science* 244, 1288-1292 (1989). The transgene is then introduced into an embryonic stem cell, where it undergoes homologous
15 recombination with an endogenous variant gene. Mice and other rodents are preferred animals. Such animals provide useful drug screening systems.

 In addition to substantially full-length polypeptides expressed by variant genes, the present invention includes biologically active fragments of the polypeptides, or analogs thereof, including organic molecules which simulate the
20 interactions of the peptides. Biologically active fragments include any portion of the full-length polypeptide which confers a biological function on the variant gene product, including ligand binding, and antibody binding. Ligand binding includes binding by nucleic acids, proteins or polypeptides, small biologically active molecules, or large cellular structures.

- 25 Polyclonal and/or monoclonal antibodies that specifically bind to variant gene products but not to corresponding prototypical gene products are also provided. Antibodies can be made by injecting mice or other animals with the variant gene product or synthetic peptide fragments thereof. Monoclonal antibodies are screened as are described, for example, in Harlow & Lane, *Antibodies, A Laboratory Manual*,
30 Cold Spring Harbor Press, New York (1988); Goding, *Monoclonal antibodies, Principles and Practice* (2d ed.) Academic Press, New York (1986). Monoclonal antibodies are tested for specific immunoreactivity with a variant gene product and

lack of immunoreactivity to the corresponding prototypical gene product. These antibodies are useful in diagnostic assays for detection of the variant form, or as an active ingredient in a pharmaceutical composition.

IV. Kits

5 The invention further provides kits comprising at least one allele-specific oligonucleotide as described above. Often, the kits contain one or more pairs of allele-specific oligonucleotides hybridizing to different forms of a polymorphism. In some kits, the allele-specific oligonucleotides are provided immobilized to a substrate. For example, the same substrate can comprise allele-specific
10 oligonucleotide probes for detecting at least 10, 100 or all of the polymorphisms shown in Figures 5A-5Q. Optional additional components of the kit include, for example, restriction enzymes, reverse-transcriptase or polymerase, the substrate nucleoside triphosphates, means used to label (for example, an avidin-enzyme conjugate and enzyme substrate and chromogen if the label is biotin), and
15 the appropriate buffers for reverse transcription, PCR, or hybridization reactions. Usually, the kit also contains instructions for carrying out the methods.

From the foregoing, it is apparent that the invention includes a number of general uses that can be expressed concisely as follows. The invention provides for the use of any of the nucleic acid segments described above in the diagnosis or
20 monitoring of diseases, such as coronary artery disease, diabetes, coagulation disorders, lipid metabolism disorders, energy metabolism disorders, diseases of the blood, blood vessels and cardiovascular system, and infection by microorganisms, as well as psychological disorders (e.g., bipolar disorder, schizophrenia). The invention further provides for the use of any of the nucleic acid segments in the
25 manufacture of a medicament for the treatment or prophylaxis of such diseases. The invention further provides for the use of any of the DNA segments as a pharmaceutical.

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled
30 in the art that various changes in form and details may be made therein without

departing from the spirit and scope of the invention as defined by the appended claims.

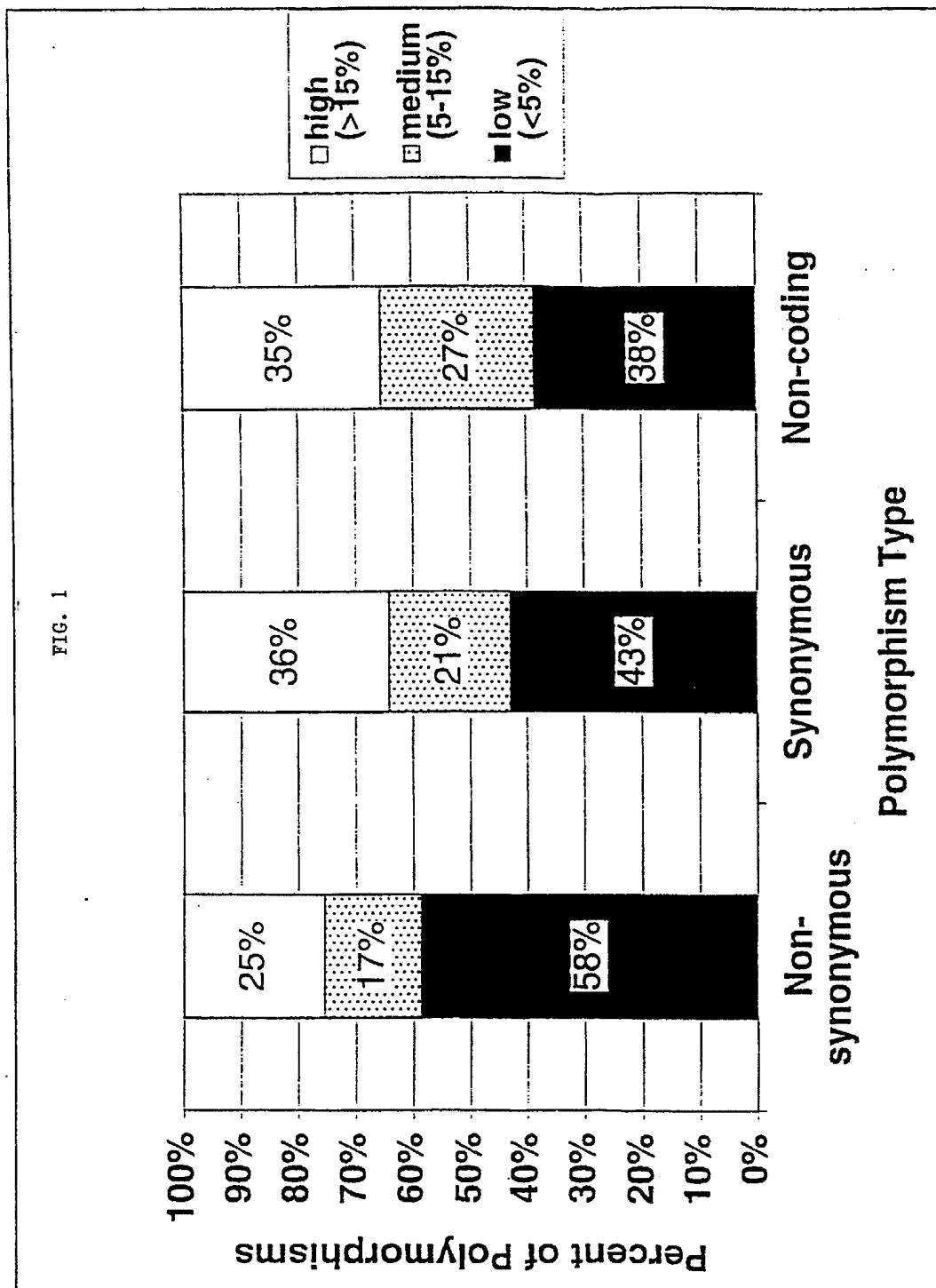
CLAIMS

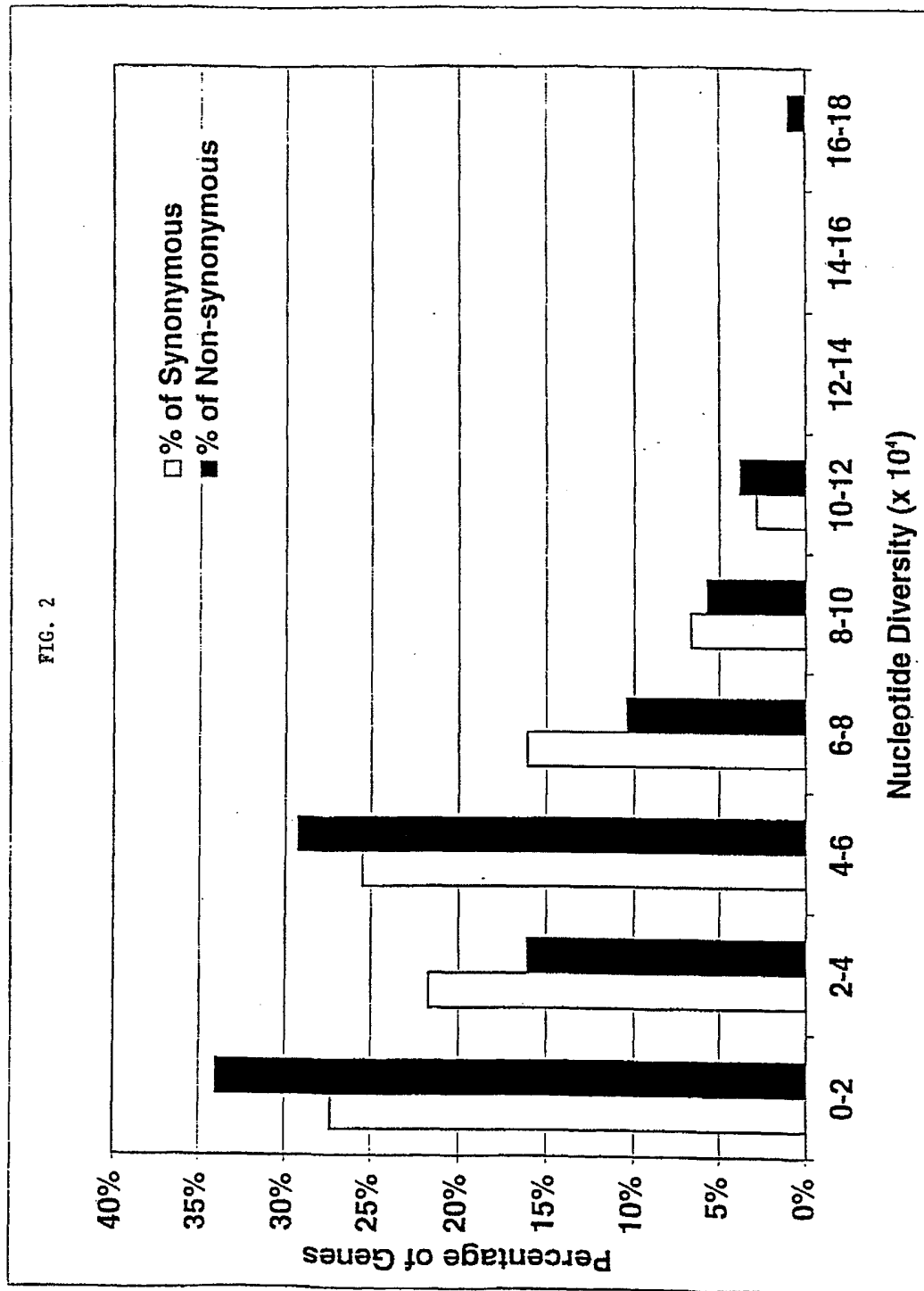
What is claimed is:

1. A nucleic acid molecule selected from the group consisting of the genes listed in Figures 5A-5QQQQQQQ, wherein said nucleic acid molecule is at least 5
5 nucleotides in length and comprises a polymorphic site identified in Figures 5A-5QQQQQQQ, wherein a nucleotide at the polymorphic site is different from a nucleotide at the polymorphic site in a corresponding reference allele.
2. A nucleic acid molecule according to Claim 1, wherein said nucleic acid molecule is at least 10 nucleotides in length.
- 10 3. A nucleic acid molecule according to Claim 1, wherein said nucleic acid molecule is at least 20 nucleotides in length.
4. A nucleic acid molecule according to Claim 1, wherein the nucleotide at the polymorphic site is the variant nucleotide for the gene listed in Figures 5A-5QQQQQQQ.
- 15 5. An allele-specific oligonucleotide that hybridizes to a portion of a gene selected from the group consisting of the genes listed in Figures 5A-5QQQQQQQ, wherein said portion is at least 5 nucleotides in length and comprises a polymorphic site identified in Figures 5A-5QQQQQQQ, wherein a nucleotide at the polymorphic site is different from a nucleotide at the
20 polymorphic site in a corresponding reference allele.
6. An allele-specific oligonucleotide according to Claim 5 that is a probe.
7. An allele-specific oligonucleotide according to Claim 5, wherein a central position of the probe aligns with the polymorphic site of the portion.

8. An allele-specific oligonucleotide according to Claim 5 that is a primer.
9. An allele-specific oligonucleotide according to Claim 8, wherein the 3' end of the primer aligns with the polymorphic site of the portion.
10. An isolated gene product encoded by a nucleic acid molecule according to
5 Claim 1.
11. A method of analyzing a nucleic acid sample, comprising obtaining the nucleic acid from an individual sample; and determining a base occupying any one of the polymorphic sites shown in Figures 5A-5Q. Q. Q. Q. Q. Q. Q. Q. Q.
- 10 12. A method according to Claim 11, wherein the nucleic acid sample is obtained from a plurality of individuals, and a base occupying one of the polymorphic positions is determined in each of the individuals, and the method further comprising testing each individual for the presence of a disease phenotype, and correlating the presence of the disease phenotype with the base.

1/178





3/178

Gene	coding bp screened	No. Synonymous polymorphisms	No. Non- synonymous polymorphisms	Non-coding bp screened	No. Non-coding polymorphisms
AADC	1229	0	2	311	0
ADORA2	332	0	1	75	0
AHC	1413	0	0	63	1
ANX3	929	2	4	725	6
APOD	570	1	3	383	1
AR	2759	3	1	300	0
AT3	1357	3	0	121	0
BDNF	744	0	1	212	0
CD36	1209	1	1	252	0
CETP	1397	4	4	299	0
CGA	349	1	0	235	0
CLanalog	1461	3	2	12	0
CNTF	603	0	1	154	0
COMT	783	2	1	241	1
CRH	51	0	0	745	3
CYP11A	1556	1	1	547	0
CYP11B1	1410	7	7	496	9
CYP11B2	1512	7	8	906	4
CYP17	1395	3	0	36	0
CYP21	1488	5	11	542	7
DBH	1266	0	2	49	0
DRD1	1341	1	0	81	0
DRD2	1032	2	0	1379	3
DRD3	719	0	1	145	0
DRD5	1408	2	1	34	0
F10	1369	3	2	416	1
F11	1878	7	4	1312	2
F13A1	2199	3	6	948	4
F13B	1952	4	6	2339	4
F2	1740	3	2	292	0
F2R	1202	2	1	13	0
F3	875	0	1	92	0
F5	6564	13	16	1542	8
F7	1262	4	2	1209	2
F9	1364	0	1	1062	2
FGA	1935	2	2	490	0
FGB	1476	7	3	1057	0
FGG	1252	0	2	1392	2
FSH	355	1	1	44	0
FSHR	1683	1	3	0	0
GABRB1	1425	5	0	804	2
GAP43	675	1	1	79	0
GH1	644	0	1	426	5
GHR	1765	1	6	391	1
GNRHR	237	0	1	513	0
GP1BA	1881	2	2	48	0
GP1BB	1238	0	0	73	0
GP5	1683	0	0	52	0
GP9	534	1	0	143	0
GRF	224	0	0	239	0
GRIN1	1681	1	0	553	0
GRL	2334	4	3	4028	5
HCF2	1500	3	3	64	1

FIG. 3A

4/178

Gene	coding bp screened	No. Synonymous polymorphisms	No. Non- synonymous polymorphisms	Non-coding bp screened	No. Non-coding polymorphisms
HMGCR	1724	0	1	12	1
HSD3B1	1122	3	2	653	1
HSD3B2	1122	1	1	723	2
HTR1A	1272	1	0	1189	1
HTR1D	1134	1	1	46	0
HTR1DB	1173	2	0	85	1
HTR1E	1098	1	1	70	0
HTR1EL	1101	1	0	46	0
HTR2A	1398	2	3	1709	9
HTR2C	1245	0	1	138	0
HTR5A	1062	2	0	34	0
HTR6	437	1	0	34	0
HTR7	1279	0	0	138	0
IGF1	630	0	0	7250	8
IGF2	546	0	0	610	1
ITGA2B	2833	4	3	707	0
ITGB3	2131	4	3	163	0
KLK2	297	0	1	279	2
LCAT	1289	1	2	90	0
LDLR	2101	7	3	38	0
LIPC	1471	4	3	754	4
LPL	409	1	1	48	0
MAOA	1032	1	0	69	0
MAOB	980	1	0	135	0
MPL	1748	1	2	903	1
NGFB	726	1	1	1186	5
NOS1	127	0	0	56	0
NT3	774	1	0	150	0
NTRK1	1961	5	2	1106	0
PACE	1500	2	0	1095	4
PAI1	1171	1	2	911	1
PAI2	1248	5	4	915	5
PC1	1881	1	3	456	1
PCI	1221	5	5	576	4
POMC	132	0	0	520	0
PRL	633	1	1	180	1
PROC	1334	3	0	114	0
PROS1	1868	1	0	557	0
PTAFR	1029	0	2	13	0
PTH	348	1	0	230	2
PTHLH	634	0	0	2342	13
SELP	2096	5	8	14	0
SHBG	1209	1	3	494	1
SLC6A1	1388	2	0	547	2
SLC6A3	1496	6	1	205	0
SLC6A4	1623	1	2	824	1
TBXA2R	1006	1	0	12	0
TBXAS1	1605	1	6	1411	1
TFPI	806	0	1	139	0
TH	965	1	1	104	0
THBD	1728	0	0	26	0
THPO	1049	0	0	632	2
VLDLR	2391	3	1	850	2
ALL GENES	135823	203	192	59552	150

FIG. 3B

5/178

Polymorphism rates for different classes of sites. Nucleotide diversity and heterozygosity (π) are expressed $\times 10^4$.

Polymorphism Type	bp screened	No. polys	Adjusted for frequency of sites*			
			Frequency (SNP/bp)	$\hat{\theta}$	π	π
Non-coding	59,552	150	1/397	4.93 ± 1.24	5.05 ± 2.40	
Coding	135,823	395	1/344	5.47 ± 1.32	5.07 ± 2.40	
synonymous		203	1/669	2.81 ± 0.68	2.98 ± 1.42	9.84 ± 2.38
non-synonymous		192	1/707	2.66 ± 0.64	2.06 ± 0.98	10.43 ± 4.97
conservative		122	1/1113	1.69 ± 0.41	1.44 ± 0.68	3.73 ± 0.90
non-conservative		70	1/1940	0.97 ± 0.23	0.63 ± 0.30	2.89 ± 1.37
four-fold degenerate sites	21,645	111	1/195	9.64 ± 2.32	9.26 ± 4.40	4.94 ± 1.19
two-fold degenerate sites	34,294	125	1/274	6.85 ± 1.65	5.33 ± 2.53	4.21 ± 1.99
non-degenerate sites	79,659	157	1/507	3.70 ± 0.89	2.52 ± 1.19	2.61 ± 0.63
Total	195,375	545	1/357	5.31 ± 1.28	5.01 ± 2.38	1.70 ± 0.81

* The number of synonymous sites was calculated as the sum of four-fold degenerate sites and half the number of two-fold degenerate sites; the number of non-synonymous sites is the sum of the non-degenerate sites and half the two-fold degenerate sites. The number of conservative and non-conservative sites is estimated as the proportion of non-synonymous sites at which a nucleotide substitution would create a conservative or non-conservative substitution, calculated as in footnote 21.

FIG. 4

6/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
AAAD04	AADC	C	T	P	L	cds	GE1048	CACACACCTGT ACAAATCCAA	CTTTACAAAGAA AGGNATCAGGC	CACACACCTGTACAAATCCAACTctgtctcttcttccaggcacactcctcagtggaagaagctt194 gggttaattggtagtgaattaaagccatcc[c/t]ctcagatggcaactctgcactcgcgtg cgtctcccttggaggaagcccttggagagagacaagcggctggcctgATCTCTTCTTTGTAAAG
AAAD05	AADC	A	G	M	V	cds	GE1048	CACACACCTGT ACAAATCCAA	CTTTACAAAGAA AGGNATCAGGC	CACACACCTGTACAAATCCAACTctgtctcttcttccaggcacactcctcagtggaagaagctt194 gggttaattggtagtgaattaaagccatccctcagatggcaactcgcct[a/g]tgcgtg cgtctcccttggaggaagcccttggagagagacaagcggctggcctgATCTCTTCTTTGTAAAG
AAAD06	AADC	G	A	V	M	cds	GE1094	CACCTGAATCAT TTTCTTTCTGC	ACACACTTACC CCAGGC	CACCTGAATCATTTTCTTTCTGAGtttcaactctgcagagccacacacatgaacgaagt1267 aatctccgaaggagagggaggagatgggtgattac[g/a]tggcccacacacatgaagacattga ggggcccgaggtctacctgcgctggagcgggtactcgcgtgcgcctgctacctcctgcctgcctcc ctcaggagccagacagcttttgaggacatcatcaacgagcttgagaagataaatcatSCCTGGGGTA AGTGCT
AAAD07	AADC	C	T	D	D	cds	GE1263	CCCTGTGTACT GCTGACCCC	CACCTCTCCCC CTTCTC	CCCTGTGTACTGCTGACCCCAaatttatgcaatgctngtccctctggataataatccactccctcc ccacgacagcggcgaagcccaagctcacagagctggagacgtgatgatga[c/t]tggctc gggaagatgctggagactaccgaagcatttttgatggagaagctgGAGAGAGGGGAGAGGTG
AAAD01	AADC	C	T	R	W	cds	GE1094	CACCTGAATCAT TTTCTTTCTGC	ACACACTTACC CCAGGC	CACCTGAATCATTTTCTTTCTGAGtttcaactctgcagagccacacacatgaacgaagt1267 aatctccgaaggagagggaggagatgggtgattacgctggcgaacacatgagagagcattgagga cgccaggtctacactgacgtggagccgggttaactg[c/t]ggccgtgatccctcctgcctgcctcc ctcaggagccagacagcttttgaggacatcatcaacgagcttgagaagataaatcatSCCTGGGGTA AGTGCT
AAAD02	AADC	A	G	E	G	cds	GE1094	CACCTGAATCAT TTTCTTTCTGC	ACACACTTACC CCAGGC	CACCTGAATCATTTTCTTTCTGAGtttcaactctgcagagccacacacatgaacgaagt1267 aatctccgaaggagagggaggagatgggtgattacgctggcgaacacatgagagagcattgagga cgccaggtctacactgacgtggagccgggttaactgctggcgcgtgatccctcctgcctgcctcc ggagccagacagctttg[a/g]ggacatcatcaacgagcttgagaagataaatcatSCCTGGGGTA AGTGCT
AAAD03	AADC	A	T	I	I	cds	GE1004	TCCATCTGGGACT ACTCAC	GTGCACCTACC TCCACTC	TCCATCTGGGACTCACAGattgattggaactttaggcaacaggaagacat[a/t]tggctgc acgttgatgcagctacgcagcgagctgattcaactcctgcctggagttccggcaccttctgaaatgGA GTGGAGGTAGTGTGCAC
ADORA2 ul	ADORA 2	C	T	A	V	cds	GE1141	ATGGACCGTGA GCTGGC	AAGCTGGGCA CCAAACA	ATGGACCGTGAAGTGTGCCcagcccgctcctgctgagcctgctgcgtctgtggccatgccca tcatgggctcctcgggtgtacatacagggtggagctggccattgtgtgctggccactcctgggcaat gtcctgtggctggggccgtggtcctcaacagcaacdtgagaagcagaaccaactactttgtgggt gtacttgggg[c/t]ggccacatcgacgtgggtgtgctgcacatccctcttgcacacacatc agacccgggttctgctgctgcacagctgccttcttcatgtccttgccttgcctggctccac cgagactcacttcatgctcctggccatgccatgacgcgtatcatgctcactcgcgatccgcgc tccgggtgagcagggccgggttactctgtgcaaggctgtgtgtgtGTGTGCCACAGGCTT
ADORA2 u2	ADORA 2	C	G	P	P	cds	GE1141	ATGGACCGTGA GCTGGC	AAGCTGGGCA CCAAACA	ATGGACCGTGAAGTGTGCCcagcccgctcctgctgagcctgctgcgtctgtgtggccatgcc[c /g]atcatgggtcctcgggtgtacatacaggtggagctggccattgtgtgctggccactcctggg caaigtgctgggtgtcctggggccgtggtcctcaacagcaacdtgagaagcagaaccaactcttgg tgggtcacttggggccggccacatcgacgtgggtgtgctgcacatccctcttgcacacacatc agacccgggttctgctgctgcacagctgccttcttcatgtccttgccttgcctggctccac cgagagctcacttcatgctcctggccatgccatgacgcgtatcatgctcactcgcgatccgcgc tccgggtgagcagggccgggttactctgtgcaaggctgtgtgtgtGTGTGCCACAGGCTT

FIG. 5A

[illegible]

[illegible]

FIG. 5D

[illegible]

[illegible]

FIG. 51.

[illegible]

[illegible]

[illegible]

FIG. 5M

poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/ noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
CYP11B B1	CYP11B B1	G	G	-	-	noncoding	GE577	GAATGGCCTG AATGGC	CTCCAGGGTCT CTGAGGCTG	GAATGGCCTGAATGGCCTTCAACCGATTGGCGTGAATCCAGAAAGTGTGTGTCGCCAACGCTG TGCAGAGTTCTCCCGATGGTGGATGAGTGGCGAGGAGCTTCTCCAGGCGCTGAAGAAGAG GTGCTGAGAAACCCCGGAGAGCTGACCTTGAGAGCTCCAGGCGAGCATCTTCCACTCAACCAT AGAGGTGTGGGCGCATGATGGGAGAGATCAGGCTCAGAGACCTTGAG
CYP11B B1	CYP11B B1	A	A	L	L	cds	GE617	ATGGCACTCAG GGCAAA	AGGCTCTGGG TGTGCC	ATGGCACTCAGGGCAAGAGCAGAGTGTGCAATGGAGTGGCTGGGTGTGCTCCCTGCAAAAGGCAACA GGCACTGGGCAAGAGAGCGCGCGGCTCCCAATGAGTGTGCTGCTTTGAGCGAGTGGCCAGG GTCCAGCAACAGGTGGTGGAGTGTGGAGTGTGGAGGAGCAGGTATATGAGGAGCTGCAC CTGAGTACACAGACTTCCAGGACTGATGGGCGCATTTTCAGGTAAAGCGCTCCCTGGC CCAGCTGGGAAACACCCAGAGCCCT
CYP11B B1	CYP11B B1	A	F	L	L	cds	GE577	GAATGGCCTG AATGGC	CTCCAGGGTCT CTGAGGCTG	GAATGGCCTGAATGGCCTTCAACCGATTGGCGTGAATCCAGAAAGTGTGTGTCGCCAACGCTG TGCAGAGTTCTCCCGATGGTGGATGAGTGGCGAGGACTTCAATCCAGGCGCTGAGCATCTTCCACTCAACA GAGGTGTGAGAAACCCCGGAGAGCTGACCTTGAGAGCTCCAGGCGAGCATCTTCCACTCAACA CCATAGAGGTGTGGGCGCATGGGAGATCAGGCTCAGAGACCTTGAG
CYP11B B1	CYP11B B1	G	K	R	R	cds	GE577	GAATGGCCTG AATGGC	CTCCAGGGTCT CTGAGGCTG	GAATGGCCTGAATGGCCTTCAACCGATTGGCGTGAATCCAGAAAGTGTGTGTCGCCAACGCTG TGCAGAGTTCTCCCGATGGTGGATGAGTGGCGAGGACTTCTCCAGGCGCTGAAGAAG GAGGTGTGAGAAACCCCGGAGAGCTGACCTTGAGAGCTCCAGGCGAGCATCTTCCACTCAACA CCATAGAGGTGTGGGCGCATGGGAGATCAGGCTCAGAGACCTTGAG
CYP11B B1	CYP11B B1	T	T	I	I	cds	GE625	GGAGGCAGCCA GGAGGC	GTGTCCCTTCC CCATAGCAC	GGAGGCAGGCAGGAGGCCTGGCTGTGTGCTCAGAGTGCATCTCCCGAGGAGCAGCAAC TTGGCTTTTGGAGAGCGTGGGCTGTGTGGCAGAGCGCCAGTCTGTCGAGGCTGAACT TCTCCATGCGTGGAGTGTGATCAATCCAGCTCCAGTCACTGATGATGCGCAGGAGCTG TCTCGTGAATGTCAGCCCAAGTGTGGAGAGGAGCACTTGGAGGCTGGAGTGCATCTTC AGTACGTTGAGGCGGAGGAGCGGCGATGATGGGCGAGGAGCAAC
CYP11B B1	CYP11B B1	A	A	A	A	cds	GE570	TCCAGCACCA AAGTCTGAG	GGCATCACCT CTCTGGGT	TCCAGCACCAAAAGTCTGAGGAGTGTCTCCGCTCCCGATGAGGAGCAACTGTATCCAGAAA TCTATCAGGAAGTGGCTTCAAGCGCTCAACAGTCAACAGCATGCTGGGCTGATGAGTGCCTCT GTGAATGGGAACTGTCCAGATGACATCAAGGCCAACTCTATGAACTCACTGAGGAGAGCG TGGGAGGCTGAGGCGGAGCAACCGCGCCACCCACCCAGAGGATGATGCC
CYP11B B1	CYP11B B1	A	A	-	-	noncoding	GE570	TCCAGCACCA AAGTCTGAG	GGCATCACCT CTCTGGGT	TCCAGCACCAAAAGTCTGAGGAGTGTCTCCGCTCCCGATGAGGAGCAACTGTATCCAGAAA TCTATCAGGAAGTGGCTTCAAGCGCTCAACAGTCAACAGCATGCTGGGAGAGTCTCTGTG AATGGGAGTGTGCGGAGTCAATCAAGCCAACTCTATGAACTCACTGAGGAGAGTGGGAG CAGGTCAGGCTGATGGAACCCAGCCACCCAGAGAGGATGATGCC
CYP11B B1	CYP11B B1	G	N	D	D	cds	GE582	CTCTGTGCA GGTCTG	CTTCAGCAGGG GGCCAG	CTCTGTGCAAGTCTGACCTTGACCTGTGCTCTCTGAGAGCGTGTCTTCTCTGTCTGATGAAC GCTCTTGGAGTGGCTGGAACCCATGACGTCGAGGAGGCGGCTGCGCAGGAGAGGCTGGCC GGCGAGCAGCATCAGTGAATCAATCCAGGAAGCAACAGCGAGTCCCTTGTGCTGTGCGG CCCTCAAGGAGACTTGGCGTGGTGTGGCTGAGGCTCCCTGTGGCCCTGCGCCCTCTGCTGA G
CYP11B B2	CYP11B B2	T	T	-	-	noncoding	GE1213	GAGGACTGAAG GGAGTGTG	CCACTGGGTGG TGGAGA	GAGGACTGAAGGAGTGTGGGAGGAGCAGCAGGAGGCGCGGCTGTGCTTGTCTCAGCAGTGA TCTCCCGCAGCAGCAACTTGTCTCTTCTGGAGAGCGGTGGGCTGGTGTGGCCACAGCGCC AGTCTGCGAGCTGAATCTCTCAATGCTTGGAGTGAATGATGATCAATCCCGCTCCAGTCCAT GTCTGCGCAGGAGCTGTCTGTGATCAGCCCAAGTGTGAGGAGCACTTGGAGCT GGGACTGATCTTCAGTACGTTGAGCGAGGAGCGAGTGTGATGGGAGAGGAGCAACCAT GGGCGCAATTTCTCTCTCCACCCACCCAGTGG

FIG. 50

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence	
CYP11B 2d23	CYP11B B2	A	A	-	-	noncoding	GE533	CATCCAGCTGA GGACCCCTTT	ACTGGGAGGG AGGTCTCT	CATCCAGCTGAGACCTTTC[g/a] tggatgcccccaactccaggctctatccctgtggtgtg tcttggagagagtggtgagctgagctgagctttagaactaccacatccagctgggtgag tgaagcccccacacccctcgagctGAGAACCCTCCCTCCCACT	172
CYP11B 2d24	CYP11B B2	A	G	N	S	cds	GE587	ATGCTTCCAG CACCAGAT	GGCATCACCT CTCTGGG	ATGCTTCCAGCACCAGATCTGaggggtgctccctggacaggtgacaactatccca gaaatctaccaggaactggccttca [a/g] tccctctcaacactcacagctatcgtggcagag ctcctgttgaaggaggaaactgtcactagagcccaacagcccaactctatggaactcaactgcagg gagctgtgacacaggttggagccgagccagcccaCCGAGAGGGTGTATGCC	248
CYP11B 2d25	CYP11B B2	A	C	R	R	cds	GE588	GAGTCTCTCTG TGCAAGTC	CTCCAGCAGGG GGCCAG	GAGTCTCTCTCTGCAAGGTCTGagacactgcagacatggcttctgtagacagcttcccttgcgtga tgacgtcttctgagctgggtggaaaccccgagctgcagcagatcttgcgccaggagagctggcc gcccagccagcatcagtgaaatcccccagaggaacacccagctgaccttgccttgcggggggc cctcaaggagaccttg [a/c] ggtgggtgctggatgagggctccctgtggccCTGGCCCCCTGCT GGAG	264
CYP11B 2d26	CYP11B B2	G	A	-	-	noncoding	GE610	CTGTGCTCTG CTGGG	CAGGTCCTG GGGCTG	CCTGTGCTTGTGCTGGGCGGCTcaagctctgacctggcctctgttaggaatgggctgaatgg cgcttcaacgattgcggctgaaccagatgtgctgtgcaccaaggccgtgcagaggttctctccc gatgtgatgcagtgccaggaacttctccagccctgaagaagaagtgctgcagaacgccc ggggagcctgacctggagcgtccagcccaagcttccactacacataagagggtgtggggccat gggggaag [g/a] tccAGCCCCAGAGACCTTG	292
CYP11B 2u1	CYP11B B2	A	G	K	R	cds	GE610	CTGTGCTCTG CTGGG	CAGGTCCTG GGGCTG	CCTGTGCTTGTGCTGGGCGGCTcaagctctgacctggcctctgttaggaatgggctgaatgg cgcttcaacgattgcggctgaaccagatgtgctgtgcaccaaggccgtgcagaggttctctccc gatgtgatgcagtgccaggaacttctccagccctga [a/g] gaagaaggtgtgcagaac gcccggggagcctgacctggagctccagccagcatcttccactacacataagagggtgtggg ccatgggggaaggtccAGCCCCAGAGACCTTG	292
CYP11B 2u10	CYP11B B2	G	T	R	R	cds	GE588	GAGTCTCTCTG TGCAAGTC	CTCCAGCAGGG GGCCAG	GAGTCTCTCTCTGTCACAGGTCTGagacctgcagacatggcttctgtagacagcgttcccttgcgtga tgacgtcttctgagctggctggaaaccccgagctgcagcagatctctgcgccaggagagcctggcc gcccagccagcatcagtgaaatcccccagaggaacacccagctgaccttgccttgcgtg [g/t] g cgccctcaaggagaccttgaggtgggtgctggatggggcctccctgtggccCTGGCCCCCTGCT GGAG	264
CYP11B 2u11	CYP11B B2	G	A	A	T	cds	GE637	TCCTGGGTGAG ATAAAGGATT T	AGGATCTGGG TGTTCCT	TCCTGGGTGAGATAAAGGATTGGgctgaacaggggtggggggagagcatgggaatggcactcaggg caaggcagaggtgtgctgtggcagccctggctgctccctgcaaaaggccacgggactgggcaact agagcc [g/a] ctccggcccttagagcgggtgctgcgttctgaagccatgccagatccaggca acaggtgctgaggtctgcagatctggaggagagaggttatgagcactgcacctggagatg caccagaccttccaggagctggggccatcttccaggtaaggcctccctggcctggctGGGAAAC ACCCAGATCCCT	337
CYP11B 2u12	CYP11B B2	G	A	P	P	cds	GE637	TCCTGGGTGAG ATAAAGGATT T	AGGATCTGGG TGTTCCT	TCCTGGGTGAGATAAAGGATTGGgctgaacaggggtggggggagagcatgggaatggcactcaggg caaggcagaggtgtgctgtggcagccctggctgctccctgcaaaaggccacgggactgggcaact agagcctggggcccttagagcgtgtgc [g/a] tttgaagccatgccagatccaggca acaggtgctgaggtctgcagatctggaggagagaggttatgagcactgcacctggagatg caccagaccttccaggagctggggccatcttccaggtaaggcctccctggcctggctGGGAAAC ACCCAGATCCCT	337
CYP11B 2u13	CYP11B B2	A	G	A	A	cds	GE587	ATGCTTCCAG CACCAGAT	GGCATCACCT CTCTGGG	ATGCTTCCAGCACCAGATCTGaggggtgtccctgctccctggagaggtgacaactatccca gaaatctaccaggaactggccttcaacccgctcacaactacacaggaactgtggc [a/g] gag ctcctgttgaaggagagctgtcactagagccatcagggccacactctatggaaactcaactgcagg gagctgtgacacaggttcaggccagcagccaccccaCCGAGAGGGTGTATGCC	248

FIG. 5P

[illegible]

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
CYP21u 5	CYP21	A	G	N	S	cds	GE1208	CTGGCCTGGTA TGTGGGG	AGGGGAGGGT TCGTACAG	CTGGCCTGGTAATGTGGGGGccgggggcccgcctgcccgaatgtaaaatgtgggtggaggctgggtcccccgtgc cgctgaagccctcccccaccaccctgtccaccggcccgccagatcgctctctctggagcagcagcaagaa ctccagagctctggccttcggccttcggcgtgggtggcgcgtgctgctggggcgagcctggcgccgctgg agctctctgggtgtgctgaccgaactgtgagggcttcagcgtgctggcctggcctggcgcctgg ccctctcgagccctgcccactgagctctcatctcaagtcaagcttcgaagcttcgaagctggcct gcagcccggggatggggcccccacacggcgccaga[a/g]ccagtgaatggggcagcagcat gcaagccgggttaacctgaattctctcttattgctctTGTAGGAACCCCTCCCT
CYP21u 6	CYP21	G	A	R	K	cds	GE576	CTTATCATGTT CCCACCCCTC	GGGGGCTACTG TGAGAGGC	CTTATCATGTTCCACCTTCCAGccccacactctctctcgagacaagctgggtgtcta[g/a]ga ctaccggagacctgtcttggaagactatctccctgctctgtgaagcccccacaaagactccaccgct cagccctgctgtggccttcctgtactccatcggagccagctggggagcccgctgacccagagttc lgaggtgaagctgggctctgagggccactggggccacccgggTcaagcttGGCTCTCAGTAGGCCCTC
CYP21u 7	CYP21	T	A	I	N	cds	GE1225	CCACCTTCGGGT CAGCCT	GCCAGAAAAGG AGGGAGTA	CCACCTTCGGGTACAGCTTgctctcaagatgaccccgccctccgctgcacagcgccgctgtg aaactcaactgtttctccacgctctgagcagccagccggcaccctctgtggccacttgaggagg aatctctctccactcgagcatca[t/a]ctgttaccctcactctggagacaagatcaaggt gcttcacagccctcagggcccccaccccccctccctgagctctctgttctcctgaaactgaaag TACTCCCTCCCTTTTCTGGC
CYP21u 8	CYP21	T	A	M	K	cds	GE537	GGAGGCTCCT TCCCACA	CTGTGGCCCGA GGGGAG	GAGAGGCTCCTTCCACAGctcatctcctgcctgcgcagtttctccccaatccaggtct ccggagctgaagcaggcccatagagaagagatcacatcgtggaga[t/a]gcagctgaggcag cagaagtggggactgtcgtggagggCTCCCTCGGCCACAG
CYP21u 9	CYP21	C	G	L	L	cds	GE629	TGTTGCCACTC TGTTACTCTCT C	GTTCCTGGGA GGAGCC	TGTTGCCACTCTGTACTCTCTCTCCcaggccagccgctcagccgctcttcaacctctcagg agagct[c/g]gaggcagccagctgggggacatgtgatgactacatgtcccaaggggtggcgca gcccagcatgaaagaggctctggagcagctggagggcagctgcatgtcagatggctgagtgac tctgtatcgtggactgagaccagcaaacacctctcctctggcgctgttttttcttctcacc caccctgaggtgctgctggggacaagcaagggCTCTTCCACGCAAC
DBHu1	DBH	G	T	A	S	cds	GE981	CCCACACAGC ATTCTTA	GGCGCTTACT TCTATC	CCCACACAGCATTTTACTaccagagaagccggctt[g/t]ccttgggggttccagggctct ccagatatctcgcctgggaagtttcactaccacacccactggTGTATGAGAGGTAGCGGC
DBHu2	DBH	C	T	R	C	cds	GE1034	CCCCACAGGT TCAACA	CGAGCGCACT GGAAGC	CCCCACAGGTTCAACacgagga[tgt]gtcactgcctccctcagcgctcgtgtctcagcagttca cctgttccctggaactcaac[c/t]ggagctatggaagcctgtctacagcttcgcgcc catctcatgactgcaacaagTcctcagccgttcGGTTCACAGTGGCGTGC
DBHu3	DBH	G	C	G	A	cds	GE1294	CAGTGGGACC AGAGAGC	GCAGTTCCTCC ATCGGT	CAGTGGGACACAGAGCTCaccccagcatcggccgcctcagctcgtggcgagctcccgcc cccaagatcgggagggagcctctatgTcacagcagcagtgagcctctcctgttcactcctccg ggcgcaactcagggctggctcccgTgagagccctccctatccactccctcgagccgg aggggtccctggagctctcatggaatgTcagctacacccaggggacatccattccagctactg gtgcggaggtcaagctgggctcgtttg[g/c]gatgcggagccgtggaggttgagaacg cagatatcgtggtgctctgTACCGATGGGACACTGC
DBHu4	DBH	T	C	V	A	cds	GE1034	CCCCACAGGT TCAACA	GCAGCGCACT GGAAGC	CCCCACAGGTTCAACacgagga[tgt]gtcactgcctccctcagcgctcgtgtctcagcagttca cctgt[t/c]tccctggaaacttcactcagccgttccGGTTCACAGTTCGGTGC
DBHu5	DBH	C	G	A	A	cds	GE1034	CCCCACAGGT TCAACA	GCAGCGCACT GGAAGC	CCCCACAGGTTCAACacgagga[tgt]gtcactgcctccctcagcgctcgtgtctcagcagttca cctgttccctggaaacttcactcagccgtcagctgTactgaaagc[c/g]ctgtacagcttcgggcc catctcatgactgcaacaagTcctcagccgttcGGTTCACAGTTCGGTGC
DBHu6	DBH	G	A	T	T	cds	GE966	TTTCTCAGGG AGTCTCG	CTTACCACGT GCCCACTC	TTTCTCAGGGAGTGTGctcatcacctctctgcagctacacac[g/a]gaagcgggagCTGG TCACAGTCMAAG

FIG. 5W

[illegible]

FIG. 5X

[illegible]

FIG. 5CC

[illegible]

FIG. 5DD

36/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence	
DRD5ul 1	DRD5	C	G	Q	E	cds	GE1171	CCAGCAGCAA CGGCAC	CTGGCGGATGC GGTAGAT	CCAGCAGCAAACGACACGcgctaccgggggagttcgctctataccag[c/g]agctggcgaggg ggaaacggctgggggctcggggggacccctgggcccctcacaggtggtggtcaccgctgc ctgctgacctactcatctctggaccctctgggacgctgggctggtggtgcagacatcggtgcg gagcgccacctggcgcccaaatgacaaactctctcatctgctctggccgtgacagaccttt tcgttcgctgctggctcatccctggagggagtcgcccgggtggccggttactggccctttgga cgctcatcagctggagccgctactgggcttcgacatctccaggcccttcggtacagcgcaagatgactc agcgatggccttggtcatggttggcctggatgacacttgacatctcatctctcatctccg gtccagctcaactggacagggacacggcgccctttgggggggggtggacctggcccaaacctt ggccaaactggacgcccctggggagggagactttgggagcccagcgtgaatcgagagactgtgact ccagcctgaatcgaaactcgccatctctctctgctcatcagcttctacatccccgttggcctc atgactcgtgactacacggcgtATCTACCGCATCGCCAG	754
DRD5ul 2	DRD5	T	A	L	Q	cds	GE1171	CCAGCAGCAA CGGCAC	CTGGCGGATGC GGTAGAT	CCAGCAGCAAACGACACGcgctaccgggggagttcgctctataccagcgclt/a]ggcgaggg ggaaacggctgggggctcggggggacccctgggcccctcacaggtggtggtcaccgctgc ctgctgacctactcatctctggaccctctgggacgctgggctggtggtgcagacatcggtgcg gagcgccacctggcgcccaaatgacaaactctctcatctgctctggccgtgacagaccttt tcgttcgctgctggctcatccctggagggagtcgcccgggtggccggttactggccctttgga cgctcatcagctggagccgctactgggcttcgacatctccaggcccttcggtacagcgcaagatgactc agcgatggccttggtcatggttggcctggatgacacttgacatctcatctctcatctccg gtccagctcaactggacagggacacggcgccctttgggggggggtggacctggcccaaacctt ggccaaactggacgcccctggggagggagactttgggagcccagcgtgaatcgagagactgtgact ccagcctgaatcgaaactcgccatctctctctgctcatcagcttctacatccccgttggcctc atgactcgtgactacacggcgtATCTACCGCATCGCCAG	754
DRD5ul 3	DRD5	C	G	L	V	cds	GE1171	CCAGCAGCAA CGGCAC	CTGGCGGATGC GGTAGAT	CCAGCAGCAAACGACACGcgctaccgggggagttcgctctataccagcgagctggcgaggggaa ggaaacggctgggggctcggggggacccctgggcccctcacaggtggtggtcaccgctgc ctgctgacctactcatctctggaccctctgggacgctgggctggtggtgcagacatcggtgcgagc cgcaacctggcgcccaaatgacaaactctctcatctgctct [c/g]tggccgtgacagacctt tcgttcgctgctggctcatccctggagggagtcgcccgggtggccggttactggccctttgga cgctcatcagctggagccgctactgggcttcgacatctccaggcccttcggtacagcgcaagatgactc agcgatggccttggtcatggttggcctggatgacacttgacatctcatctctcatctccg gtccagctcaactggacagggacacggcgccctttgggggggggtggacctggcccaaacctt ggccaaactggacgcccctggggagggagactttgggagcccagcgtgaatcgagagactgtgact ccagcctgaatcgaaactcgccatctctctctgctcatcagcttctacatccccgttggcctc atgactcgtgactacacggcgtATCTACCGCATCGCCAG	754

FIG. 5EE

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence	
DRD5u1 4	DRD5	C	T	A	A	cds	GE1171	CCAGGCAGCAA CGGCAC	CTGGGCGATGC GGTAGAT	CCAGGCAGCAACGGCACCGCGTACCCTGGGCGAGTTCGCTCTATACCAAGCAGCTGGCGCAGGGGAA CGCGTGGGGGCTCGGGGGGCAACGCACTGGGCGCTCAACAGTGTCAACGCTGCTGCTGC TGACCTACTCATCATCTGGACCTGTGGACCTGTGGGAAAGTCTGGTGTGCGAGCCATCTGCGGAGC CGCACTGCGCGCAACATGACCAACGCTTCACTGTCTCTGCGCGTGCAGACC [T/G]TT TCGTGGCGCTGGTTCATCGCTGGAAGCAGTCGCGAGTGGCGGTACTGGCCCTTGGG GGTTCTGCACTGTGGTCTGACATCTGCTCCACTGCTCCACTGCTCCACTGCTCCACTGCT CGTCAAGCTGGACGCTACTGGGCACTCCAGGCTTCGCTCAACGCAAGATGACTC AGCGATGGCTGGTCACTGGTGGCGCATGGACCTTGCATCCCTCATCTCTCATTCG GCCAGCTCACTGGCAGGACAGGCGCTCTGGGCGGCTGGGAGCTGGAGCGGAGTGAATGCGAGAACTGACT GGCACTGAACTGAACTCACTGCTCTCTCTGCTCATCAGCTTCTACATCCCTGCTGCTGCTC CAGCTGAACTGAACTCACTGCTCTCTCTGCTCATCAGCTTCTACATCCCTGCTGCTGCTC ATGATGCTGAACTCACTGCTCTCTCTGCTCATCAGCTTCTACATCCCTGCTGCTGCTC	754
DRD5u1 5	DRD5	T	G	L	R	cds	GE1171	CCAGGCAGCAA CGGCAC	CTGGGCGATGC GGTAGAT	CCAGGCAGCAACGGCACCGCGTACCCTGGGCGAGTTCGCTCTATACCAAGCAGCTGGCGCAGGGGAA CGCGTGGGGGCTCGGGGGGCAACGCACTGGGCGCTCAACAGTGTCAACGCTGCTGCTGC TGACCTACTCATCATCTGGACCTGTGGACCTGTGGGAAAGTCTGGTGTGCGAGCCATCTGCGGAGC CGCACTGCGCGCAACATGACCAACGCTTCACTGTCTCTGCGCGTGCAGACC [T/G]TT TCGTGGCGCTGGTTCATCGCTGGAAGCAGTCGCGAGTGGCGGTACTGGCCCTTGGG GGTTCTGCACTGTGGTCTGACATCTGCTCCACTGCTCCACTGCTCCACTGCTCCACTGCT CGTCAAGCTGGACGCTACTGGGCACTCCAGGCTTCGCTCAACGCAAGATGACTC AGCGATGGCTGGTCACTGGTGGCGCATGGACCTTGCATCCCTCATCTCTCATTCG GCCAGCTCACTGGCAGGACAGGCGCTCTGGGCGGCTGGGAGCTGGAGCGGAGTGAATGCGAGAACTGACT GGCACTGAACTGAACTCACTGCTCTCTCTGCTCATCAGCTTCTACATCCCTGCTGCTGCTC CAGCTGAACTGAACTCACTGCTCTCTCTGCTCATCAGCTTCTACATCCCTGCTGCTGCTC ATGATGCTGAACTCACTGCTCTCTCTGCTCATCAGCTTCTACATCCCTGCTGCTGCTC	754
DRD5u1 6	DRD5	T	C	T	T	cds	GE1171	CCAGGCAGCAA CGGCAC	CTGGGCGATGC GGTAGAT	CCAGGCAGCAACGGCACCGCGTACCCTGGGCGAGTTCGCTCTATACCAAGCAGCTGGCGCAGGGGAA CGCGTGGGGGCTCGGGGGGCAACGCACTGGGCGCTCAACAGTGTCAACGCTGCTGCTGC TGACCTACTCATCATCTGGACCTGTGGACCTGTGGGAAAGTCTGGTGTGCGAGCCATCTGCGGAGC CGCACTGCGCGCAACATGACCAACGCTTCACTGTCTCTGCGCGTGCAGACC [T/G]TT TCGTGGCGCTGGTTCATCGCTGGAAGCAGTCGCGAGTGGCGGTACTGGCCCTTGGG GGTTCTGCACTGTGGTCTGACATCTGCTCCACTGCTCCACTGCTCCACTGCTCCACTGCT CGTCAAGCTGGACGCTACTGGGCACTCCAGGCTTCGCTCAACGCAAGATGACTC AGCGATGGCTGGTCACTGGTGGCGCATGGACCTTGCATCCCTCATCTCTCATTCG GCCAGCTCACTGGCAGGACAGGCGCTCTGGGCGGCTGGGAGCTGGAGCGGAGTGAATGCGAGAACTGACT GGCACTGAACTGAACTCACTGCTCTCTCTGCTCATCAGCTTCTACATCCCTGCTGCTGCTC CAGCTGAACTGAACTCACTGCTCTCTCTGCTCATCAGCTTCTACATCCCTGCTGCTGCTC ATGATGCTGAACTCACTGCTCTCTCTGCTCATCAGCTTCTACATCCCTGCTGCTGCTC	754

FIG. 5FF

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
DRD5u2	DRD5	G	C	C	S	cds	GE1171	CCAGCAGCAACCGGCAC	CTGGCGCATGC GTTAGAT	CCAGCAGCAGCAACCGGCACCGcgcgTaccgcgggcagcttgctctataccagcagctgcgcaggggaa cgccgtggggggctcgcgcgggggcacgcgcacatggggccctcacaggtgtgaccgcctccctgcg tgacctactcatctatgtggacctgtgggaacgtctgtggttg/ctcgagcactcgtgcg ggcgccacccctgcgcgaacatgacaaacgtctctatctgtctctgtgcggctgacagaccttt tcgtgcgtgctgcgtacatgcctgcgaagcagctgcgcggagtcggcgttactggccctttggga ggcttctgcgcgttggttggtctgcacatcatgtctccactgcctccatctgaaacctgg cgtctacagcgtggacgcctactgggcaatctccagcccttcgctacaaagcgaatgactc agcgaatggccttggctcatgctgcggctggcagaccttgccatccctcatctctattccg gtccagctcaactggcacaggaacacagggccctcttggggcgggctggacctgccaaacaaact ggccaaactggacccctggggaggaggaccttttgggaagcccgactgaaagcgaggaactgtgact ccctgtgaatcgaaactcgcctctctctctgcgtctacagctctctacatcccccgttggccatc atgatgtgactctacgcgcgtCTACCGCATCGCCACAG
DRD5u2	DRD5	G	A	R	H	cds	GE1171	CCAGCAGCAACCGGCAC	CTGGCGCATGC GTTAGAT	CCAGCAGCAGCAACCGGCACCGcgcgTaccgcgggcagcttgctctataccagcagctgcgcaggggaa cgccgtggggggctcgcgcgggggcacgcgcacatggggccctcacaggtgtgaccgcctccctgcg tgacctactcatctatgtggacctgtgggaacgtctgtggttg/ctcgagcactcgtgcg ggccacctcgcgcgaacatgacaaacgtctctatctgtctctgcgcgttcagacactttcgt ggcgctgggtggtcatgcccctggaagcagctgcgcgggtggccggttactggcccttggagcgt tttgagcgtctgggtggcctcgacaatcatgtgtctcaactgcctccatccctgacacgtgcgctc atcagcgtggcgcctactggcctctccagggcccttcggctacaagcgagatgactcagcg ctggcccttggctcatggtcggccctggcatggaccttgctcatctcatctctcatctccggtcc agctcaactggcacagggacggccctcttggggggcggtggacctgccaaacaaactggcc aaactggacccctggagaggagaccttttggagaccgcagctggaatcgagagactgtgactccag cctggaatcgaaactcagcactctctctcgtctcagctgtctacatcccccgttggccatcatga tcgtgacctacagc/g/a/ctCTACCGCATCGCCACAG
DRD5u3	DRD5	C	G	N	K	cds	GE1171	CCAGCAGCAACCGGCAC	CTGGCGCATGC GTTAGAT	CCAGCAGCAGCAACCGGCACCGcgcgTaccgcgggcagcttgctctataccagcagctgcgcaggggaa cgccgtggggggctcgcgcgggggcacgcgcacatggggccctcacaggtgtgaccgcctccctgcg tgacctactcatctatgtggacctgtgggaacgtctgtggttg/ctcgagcactcgtgcg ggccacctcgcgcgaacatgacaaacgtctctatctgtctctgcgcgttgcagaccttt tcgtggcgtctgtgctacgtccctgcgaagcagctgcgcgggtggccgttactggccctttgga ggcttctgcgcgtctgggtggccttcgacaatcatgtgtctcactgcctccatccatgaacctgtg cgtctacagcgtggacgcctactggggcactccagggcccttcgctacaagcgaagatgactc agcgaatggccttggctcatggtcggcctggcagaccttgctcatccctctctctctatccg gtccagctcaactggcacagggacggccctcttggggcgggctggacctgccaaacaaact ggccaaactggacccctggggaggaggaccttttgggaagcccgactgaaagcgaggaactgtgact ccagctgaatcgaaactcgcctctctctcgtctcagctcagctctacatcccccgttggccatc atgatgtgactctacgcgcgtCTACCGCATCGCCACAG

FIG. 5HH

40/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence
DRD5u4	DRD5	C	T	F	F	cds	GE1171	CCAGGCAGCAA CGGCAC	CTGGGCGATGC GGTAGAT	CCAGGCAGCAACCGCACCGCTACCCGGGGCAGTTCGCTATACCAGCAGCTGGCGAGGGGAA 754 CGCCGTGGGGGCTGGCGGGGGGCGCCAGCCAGTGGTCCACAGTGGTCCAGCGCTGCTGC TGACCTACTCATCATCTGGACCTCTGGGCAACGCTGGTGGTGGCGAGCCATCGTGGGAGC CGCCACTGGCGCCCAACATGACCAAGCTCTTCACTGCTGCTGGCGTTCAGACCTTTCTGT GGCGTGTGGTCACTCCCTGGAAGGAGTGGCGAGTGGCGGTTACTGGCCCTTGGAGCGT TCTGCGAGTCTGGTGGCTTCGACATCATGTCTCACTGC C/A TCCATCTGAACCTGTG CGTCAACGCTGGACCGCTACTGGCCATCTCAGCCCTTCGCTACAAGCGCAAGATGACTC AGCGATGGCTTGGTCACTGGTGGCTGGCTGGACCTTGTCCATCTCTCTCATCTCCG GTCCAGTCAACTGGCACAGGACCGGCTCTTGGGGCGGGGTGACCTGCCAACACCT GGCAACTGGACCGCTGGGAGGAGTCTTGGAGCCGAGTGAATGCAGAGAACTGTGACT CCAGCTGAATCGAACCTACGCCATCTCTCTCTGCTCATCAGCTTCTACATCCCGTGGCATC ATGATCTGACTACACGGGAGTCTACCGCATCCGCCAG
DRD5u5	DRD5	C	A	A	A	cds	GE1171	CCAGGCAGCAA CGGCAC	CTGGGCGATGC GGTAGAT	CCAGGCAGCAACCGCACCGCTACCCGGGGCAGTTCGCTATACCAGCAGCTGGCGAGGGGAA 754 CGCCGTGGGGGCTGGCGGGGGGCGCCAGCTGGGCGCCCTCAGTGGTCCACAGTGGTCCAGCGCTGCTGC TGACCTACTCATCATCTGGACCTCTGGGCAACGCTGGTGGTGGCGAGCCATCGTGGGAGC CGCCACTGGCGCCCAACATGACCAAGCTCTTCACTGCTGCTGGCGTTCAGACCTTTCTGT GGCGTGTGGTCACTCCCTGGAAGGAGTGGCGAGTGGCGGTTACTGGCCCTTGGAGCGT TCTGCGAGTCTGGTGGCTTCGACATCATGTCTCACTGC C/A TCCATCTGAACCTGTG CGTCAACGCTGGACCGCTACTGGCCATCTCAGCCCTTCGCTACAAGCGCAAGATGACTC AGCGATGGCTTGGTCACTGGTGGCTGGCTGGACCTTGTCCATCTCTCTCATCTCCG GTCCAGTCAACTGGCACAGGACCGGCTCTTGGGGCGGGGTGACCTGCCAACACCT GGCAACTGGACCGCTGGGAGGAGTCTTGGAGCCGAGTGAATGCAGAGAACTGTGACT CCAGCTGAATCGAACCTACGCCATCTCTCTCTGCTCATCAGCTTCTACATCCCGTGGCATC ATGATCTGACTACACGGGAGTCTACCGCATCCGCCAG
DRD5u6	DRD5	C	G	V	V	cds	GE1171	CCAGGCAGCAA CGGCAC	CTGGGCGATGC GGTAGAT	CCAGGCAGCAACCGCACCGCTACCCGGGGCAGTTCGCTATACCAGCAGCTGGCGAGGGGAA 754 CGCCGTGGGGGCTGGCGGGGGGCGCCAGCTGGGCGCCCTCAGTGGTCCACAGTGGTCCAGCGCTGCTGC TGACCTACTCATCATCTGGACCTCTGGGCAACGCTGGTGGTGGCGAGCCATCGTGGGAGC CGCCACTGGCGCCCAACATGACCAAGCTCTTCACTGCTGCTGGCGTTCAGACCTTTCTGT GGCGTGTGGTCACTCCCTGGAAGGAGTGGCGAGTGGCGGTTACTGGCCCTTGGAGCGT TCTGCGAGTCTGGTGGCTTCGACATCATGTCTCACTGCCTCACTGCTCACTGAACCTGTGCT C/G ATCAGCGTGGACCGCTACTGGCCATCTCAGCCCTTCGCTACAAGCGCAAGATGACTC AGCGATGGCTTGGTCACTGGTGGCTGGCTGGACCTTGTCCATCTCTCTCATCTCCG GTCCAGTCAACTGGCACAGGACCGGCTCTTGGGGCGGGGTGACCTGCCAACACCT GGCAACTGGACCGCTGGGAGGAGTCTTGGAGCCGAGTGAATGCAGAGAACTGTGACT CCAGCTGAATCGAACCTACGCCATCTCTCTCTGCTCATCAGCTTCTACATCCCGTGGCATC ATGATCTGACTACACGGGAGTCTACCGCATCCGCCAG

FIG. 5II

[illegible]

FIG. 5LL

FIG. 500

FIG. 5SS

[illegible]

FIG. 5T

FIG. 5V.V

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Assay Sequence
F5d47	F5	G	A	-	-	noncoding	GE949	TCCCTAATCAT GGAGTTTACT TT	GCACACTGTA GGGGTA	TCCTTAACATGGAGTTTACTTTAT (g/a)tatatatatttctaaagcaaaattattcattttt ttctgttttctctagctagagagcaatggagctaaagctgtatcatatctctgattcacaca ggtcagaagcttcagagtttttgggtgaagttgagcacatggtgtgtaattggtTACCCCTACAGT TTGTC
F5d48	F5	A	G	-	-	noncoding	GE316	GCAAGGTTTT AACATCTTCTT T	GCACAGTCTTC AGATTGCTTT	GCAAGGTTTACATCTTCTTctctctgttttttctcattttc(a/g)taggcaaacacaa taagcatggctagaaattgctctactctcaagatacaagataaacggcaattataacacagggt gaaattctctctctctgaattgatatgaagagctataccatccctacagtgagcagggagtg gaatggaaacacatacaggtgaaatctctcaaggcagagtagagtggaatctctgggcaaga GAAGCATCTGAAGACTGTGC
F5d49	F5	C	G	H	Q	cds	GE387	GTGCCCCAGA GGACACTA	TCAATGGCTG AGTTCTGGAG	GTGCCCCAGAGCACTATCAaaacttcccatctcaagacctgatacaatgcactctacttc agaacccagtc(a/c/g)agatctctctctccagagctcagtgaaagcttgagatgacccaagt cacagctctctccacagataaagtcaaatgctctctccagacctctccagacatgagctggcagac agctatctctccagacctccagcaggtgacctctctccagacctccagcagcaaaacctctctc cagacctccacacagactctctccagacctctctccagacctctccagacctccagacctctctc cagctgcccattctccagacctcagccatacaaacctttctccagacctcagccatacaacct ttcttagacctcagccagacaacctctctccagacctcagtcagacaacctttctccagccc tcggctcagtgcccctttctccagacctcagccatacaaacctttcttagacctcagccagaca aacctctCTCCAGACTCGCCATATGA
F5d50	F5	A	G	T	T	cds	GE389	GAAAGGTAGCT ATGAAATATC CAAGA	GGCTTTGANT GGGGAATGT	GAAAGGTAGCTATGAAATATCCAGATactaatgaagacacagctgttaacaatttgctgataca ggccccagaatgctctcagctgctgggggagaaagcacccctcttgcaacagcctggaagcag agtggccacccaaagtctctagtggttagacataaaatctcaagtaagcaggtggaggaa ggatgactgaagaaagccagttctcatttaagcacagcaaaagaaagagagagacacac acctgctcttctctccagacctctccacctctcaagtgaaagctcaaacacatttca gaagagacttaagcatctgtgtgtgtctcaataatccaatgaacatctctccacacagact caatcagac(a/g)ttggctctatgattttggctgagtagctctcctgacctataatcag aatctcacaatgaactggtcaggaagctctcctcaggtctttatcagacagtgccccca ggaacactatacaACTTCCCTTCAGACC
F5d51	F5	A	C	E	A	cds	GE496	CTTTCTCCAGA CCTTGGTGAG	AAGGAGAAAC TGSCCAAC	CTTTCTCCAGACTTGGTGACagatacttctcccaactttgttcagatgtccctttccccaga ctcagcaggtacacctctctccagacatcagtaacacacctctctccggatctcagccaga tatcactctccacagaccttgatcagatacttataacctctgaaatagctagctattgttttt caagaatttaagtagcttttctcttccagaccttggtcagatggccatctctcactctccac ttcgaatgactttctatacaaggaaattcaatccctggttctatgctggctcctcagcaagatg gtacagattacattgagatctccaaaggag(a/c)ggctcagagcgtgagatgactatgc tgaattgattgtgctctatgagtcacctcaaaactgattaggaacacatacaactctct ccagagactctgacaacattgagcatggtacctccgagcaaatggaaacagagaattat tcaattgtctgaagaaatctctgggattattcagaatttgacaaagGTTTGGCCAGTTTTC TCTT

FIG. 5WW

FIG. 5YY

[illegible]

FIG. 5ZZ

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/ noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
P5u37	P5	C	T	P	S	cds	GE391	CCCTTCTCAC CAACAAGC	AGCAGGTGAGG CATTTCTGG	CCCTTCTCACCAACAAAGCccaccagctggttcc[c/t]cactgagacacacctcatgtggcaagaa ctcagttctcaattcttccacagcagagcattccagccattatcttgaagaccctatagaggatc ctctacagccagatgtccacagggatcagctctactcttacttggctggagatcagagatcga gaacatgtcaagcgtlaaggaccccaagtagaagagatcaagcagcaaacacaggtttctcttg tgagaaattactagcacaataaagttgggagacacctcaagcagacactggcttctctccgaa tgagcccttggtggagaccttcttccagcagacactggttctctccacacagctggccctgggag gaacctctagtatgttactcttaaaacaagtaactctaatgaattttgattggagagatg gcatttggcttctggaagaggtagctagaataatccaagatactgaatgatgaagacacagctgtta acaattggctgctcagcccCCAGAAATGCTTCACGTGCT
P5u38	P5	G	C	E	D	cds	GE362	GCTATCCACAGA TTTGAAGATGG T	TTTGTCCCATG ACAGAACTCC	GCTATCCACAGATTGAAGTGGTgaagacactctcctgaagaaactgccacactgctctggagct gttactctccagttgcttcttacttgaccacacatctctctggaa[g/c]aagatggagacg ctggctccagcagagataacacattatgaattggatctcagtgagacagtggaaccacacct gatgacctccactcctcacacatctatctccctgaataatctgatctcaggtattccaactc gggctgattgggcccctgtctatctgtaaaaaaggtgaagaaacccccacacaaagattccaaca actaaatgttggaaatgctcagggatttctgTCATATGGACAA
P5u39	P5	G	A	E	K	cds	GE391	CCCTTCTCAC CAACAAGC	AGCAGGTGAGG CATTTCTGG	CCCTTCTCACCAACAAAGCccaccagctggttctccctcagacacctcatgtggcaagaactca ggttccaattctccacagcagatctccagccattcttgaacacctatagagatctctct acagccagatgtccacagggatcagctctactcttcttctggtggagaatcagagatccaagaac atgctaagcgtlaaggaccccaagtagaagagagatcaagcagaagacagaggttctctcggatg aaattactagcacaataaagttgggagacacctaaagcagaacactggttctctccggaaatgag ggccctgggagaccttctcagcagacactggttctctccagaatgagccctgg[g/a]ag gaacctctagtatgttactcttaaaacaagtaactctatcgaattttgattggagagatg gcatttggcttctggaagaggtagctagaataatccaagatactgaatcagacagcagctgtta acaattggctgctcagcccCCAGAAATGCTTCACGTGCT
P5u4	P5	T	C	M	T	cds	GE172	TTTAAAGAAAT ACAGGTCTCAG CAT	TTTCTCCCATG ATTCTGTATT GT	TTTAAAGAAATACAGGTCTCAGCATTTTgaataattcttcaacccaacttggaaacattataaga aagttatctacacacagctacagagatgagcttcttcccaacaacactcagtgaaatcccaat[a/c] gaagaaatgggatttgggtctctatctcagggccgggtcagagacacactcaaatgaatga acAAATATACAGAAATCATGCGAATA
P5u40	P5	C	G	P	A	cds	GE391	CCCTTCTCAC CAACAAGC	AGCAGGTGAGG CATTTCTGG	CCCTTCTCACCAACAAAGCccaccagctggttctccctcagacacctcatgtggcaagaactca ggttccaattctccacagcagatctccagccattcttgaacacctatagagatctctct acagccagatgtccagggatcagctctacttctcactgttggtagaattcagaagtcgaagac atgctaagcgtlaaggaa[c/g]ccaagtagaagagatcaagcagacacaggttctctcttg gatgaattactagcacaataaagttgggagacacctcaagcagaacactggttctctccgaa tgagcccttggggaccttctcagcagaacactggttctcttctccagaatgagccctgggag gaacctctagtatgttactcttaaaacaagtaactctatgaattttgattgggagatg gcatttggcttctggaagaggtagctagaataatccaagatactgaatcagacagcagctgtta acaattggctgctcagcccCCAGAAATGCTTCACGTGCT
P5u41	P5	C	G	L	V	cds	GE915	TGATATACAG A AGAGCAAGAA	ACTGTGACCA GTGTGATTTA	TGATATACAGAGAGCAAGGAAATctcttggaagggcaatacaatttactctgtttttccaggg aaaagatattcactcaggttgaaggtcccc[c/g]taactctgcanaaaggaata-taca taaggacagcaactgctatggacagagagaatttctctattatctatctcagcttggatgaa agaagctggtactatgaagaagctccagagttcttgagactcaactctcagcaaatgaa aaatcccatggtttccaggtatttctcctgacttgaatctaatctcctaaTTAAATCACTGTG GCTTCAAGT

FIG. 5DD

FIG. 5GGG

FIG. 5HHH

FIG. 5III

FIG. 5JJJ

FIG. 5LLL

FIG. 5NNN

FIG. 5000

FIG. 5PPP

74/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
FGBu9	FGB	C	T	Y	Y	cds	GE392	AAGGAAGAAA GGCAGTTT	CCAGCAAGTG GTAGCTATTAA A	AAGGAAGAAAAGCGAGTTT Ttagtttcccaaaattttattttggtgagagattttattttggttt ttctttaggtgaatttggcttggaaattttaaatttagcagatttaccagattggagcccca gaacttttagaagaatggagagactgaaaggagacaagaagctccatcattgagagattcac tgacagaatgaagccaaatcaccagatctcagtgaaacaata(c/t)lagagaaacagccgt aatgcttccatggatggagcatcctcagctggtgggaaacacagaccatgccattccaaacgg catgtttccagcagctatgacagagacaatggagctggtatggtgagcactcttctctctgc tttaaaatcacactaatatcatctactcagaatcattaaacaatttttTAATAGCTACCACTTC TGGG
FGGd3	FGG	A	G	M	V	cds	GE337	TTCGTAATAGA CAGCTCTTCAT AGACT	GCAGTTAAFTT TCTACAAATCA TCC	TTCGTAATAGACAGCTCTTCATAGACTTgcagaggttaaaagattccagaataatgatatgata tctacagactgttttaggtggcacttactcaaaagcatctactcctaattgggttaataatggca ttatttggccacttggaaacccgggtgggtattccatgaagaacacccact(a/g)lgaagataat cccatccaagagactcaaatggagagagacacccacccctggggggagccaaacaggtca gaccagagcaccctgggaacacagatatgactcactttaccctgagggatgatttGTAGAAATTT AACTGC
FGGd4	FGG	A	G	-	-	noncoding	GE349	AAATATCTTAG CAGTTTCCAAA GAAAA	TGGTATGACAC TTTCTAAACTA TTC	AAATATCTTAGCAGTTTCCAAAGAAAAtataaaattactctctctgaaaggatactatttttgt ctctctatttttgttatcttattgtttgtttgtttgtagatatattgcaggaataataataatcaata atcaaaagattgttaacctgaagaagagagagagccagcttgaagcacagtcgccaggaaccttgc aaagacacgggtgcanaatccatgatatcactgggaagaggttaactgataagaggttatatttgggattta ggttcatcaagctaatgtatgaaggagagagagctgactgg(a/g)laagtataggaATTAGTTT AGAAAGTGGCTACCA
FGGd5	FGG	G	A	G	R	cds	GE360	TGCTGATGTGA AAAGTAAGAAA AT	CAAGGTGCTTA GAAAGTATCT GC	TGCTGATGTGAAGTAAGAAAAtatcttggaaatgaatagtttactacatgttanaagcta tttttcaaggtggcagctcttactctgcatttcaaacacacagtaaaagctgactctctctct agattgtcaagcattgcctaataagggagatcaaacagagc(g/a)ggcttactttttaaactct ctgaagctaacacagcaattctttagtctactgtgaatcgatgggtctctggaatgagtgagctct gttcagaaggtaatttttcccccacatgtgtatttaataatttccataattgtttctgcata TgGCAGATACTTTTCTAAGCACCTTG
FGGd6	FGG	T	A	-	-	noncoding	GE372	GAACAGTGCT CTGTATTTTG AC	CCATTGCTAT TGATAGTTGGA AAG	GAACAGTGCTGTGTATTTTGTGCAaaatgttgacagcattctctttaca(c/a)gcattgatatag tctattttctctcttctctctctgcaaatgtgaattagagacttgatggcagtcagatttcaag aaaactggattcaataataaagaaggatttggacatctctctctactggcacaacagattttg gctgggaatgagagatctatttgaagaacacacagctctgccatcccatatgcatttaaggtgg aaactgaagactggatggcgaacacaggtactgttttgaatgacttccaaacttttctatgtaa agattgcccggatgtgcaCTTCCAACTATCAATAGCAATGG
FGGu1	FGG	A	T	Y	F	cds	GE404	CATCTTACGAA AGAGGAATTT T	TCCACTTCCAG TTTCAAGAAC T	CATCTTACGAAGAGGGAACCTTctgagatccctggagggttcagcatgtgatgttatttcc ttcttctcagttactgcagact(a/z) tgcctatttcaagtgaggactgaactgaactgaactacg cctaactatgctctactcgtgtgggggagtggtgagatgcttttgaaggcttgaatttggcg atgatcctagtgaacaagtttttcaatcccatatggcatggcttcaactctgggacaatgac aatgataagtttgaaggcaactgtctgaacagagtgatctggttgggtgagatgaacaagtgtca cgtggccatctcaatggagtttatcccaaggtatgttttctcttcttagattccaagttaag tatagtgatactatttcaaaaaataataatgagatgaagaataatgaagaataacttataa agatagagggttttatactatgcttttatttcaactaAGTTCTTTGAAACTGGAAGTGGGA

FIG. 5QQQ

FIG. 5RARR

Poly ID	Gene	ref NT	alt NT	ref AA	alt AA	coding/ noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
FSHRu4	FSHR	T	G	F	V	cds	GE667	GCTCTGAGCTT CATCCAA	AGCAAAAATCC AGCCCATCA	GCTCTGAGCTTCAATCAAAATTGCAAGCAAGCAAGTGGATATGACTCAG actaggggtcagagatcctctctggcagaagacaatagtcctcagctacacgagagga [c/g]ttg acatgcgtacacttgaatttgactatgcaatgaagtggttgagcgtgcctgtccctc aagccagatgcatcaacccatgaagatcatcatgggtgtaacaacatcctcagatcctgatatg gtttatcagcatcctggccatcactgggaacatcatagctagtgatcctcactaccagccaat ataaactcagctcccccaggttccttatgtgcaacctggcctttgtctgcatctgcatggaaatc taactgtctcatgtcatcagttgatatccataccaagagccaatatcaacattgcatcattga ctggcaactggggcaggtgtgctgtcgtggtttttcactgtctttgccagtgagctgtcag tctacactcagacgtatcaccttggaaagatggcatcacatcacatcagcatgccatcgctggac tgcgaagtgagctccgcctcgtccagtgctcatgggtgtagtggcctggatgatttggct
FSHRu5	FSHR	G	A	T	T	cds	GE667	GCTCTGAGCTT CATCCAA	AGCAAAAATCC AGCCCATCA	GCTCTGAGCTTCAATCAAAATTGCAAGCAAGCAAGTGGATATGACTCAG actaggggtcagagatcctctctggcagaagacaatagtcctcagctacacgagagatttgcacat gacig/alitacactgagtttgactatgacttatgcaatgaagtggttgagcgtgcctgtccctc aagccagatgcatcaacccatgtgagatatacatgggtgtaacaacatcctcagatcctgatatg gtttatcagcatcctggccatcactgggaacatcatagctagtgatcctcactaccagccaat ataaactcagctcccccaggttccttatgtgcaacctggcctttgtctgcatctgcatggaaatc taactgtctcatgtcatcagttgatatccataccaagagccaatatcaacattgcatcattga ctggcaactggggcaggtgtgctgtcgtggtttttcactgtctttgccagtgagctgtcag tctacactcagacgtatcaccttggaaagatggcatcacatcacatcagcatgccatcgctggac tgcgaagtgagctccgcctcgtccagtgctcatgggtgtagtggcctggatgatttggct
FSHRu6	FSHR	A	C	I	L	cds	GE667	GCTCTGAGCTT CATCCAA	AGCAAAAATCC AGCCCATCA	GCTCTGAGCTTCAATCAAAATTGCAAGCAAGCAAGTGGATATGACTCAG actaggggtcagagatcctctctggcagaagacaatagtcctcagctacacgagagatttgcacat gacgtacactgagtttgactatgacttatgcaatgaagtggttgagcgtgcctgtccctc cagatgcatcaacccatgtgagatatacatgggtgtaacaacatcctcagatcctgatatggttt atacagatcctggcc [a/c]itcactgggaacatcatagctagtgatcctcactaccagccaat ataaactcagctcccccaggttccttatgtgcaacctggcctttgtctgcatctgcatggaaatc taactgtctcatgtcatcagttgatatccataccaagagccaatatcaacattgcatcattga ctggcaactggggcaggtgtgctgtcgtggtttttcactgtctttgccagtgagctgtcag tctacactcagacgtatcaccttggaaagatggcatcacatcacatcagcatgccatcgctggac tgcgaagtgagctccgcctcgtccagtgctcatgggtgtagtggcctggatgatttggct
FSHRu7	FSHR	A	G	T	A	cds	GE667	GCTCTGAGCTT CATCCAA	AGCAAAAATCC AGCCCATCA	GCTCTGAGCTTCAATCAAAATTGCAAGCAAGCAAGTGGATATGACTCAG actaggggtcagagatcctctctggcagaagacaatagtcctcagctacacgagagatttgcacat gacgtacactgagtttgactatgacttatgcaatgaagtggttgagcgtgcctgtccctc cagatgcatcaacccatgtgagatatacatgggtgtaacaacatcctcagatcctgatatggttt atacagatcctggcc [a/g]itcgggaacatcatagctagtgatcctcactaccagccaat ataaactcagctcccccaggttccttatgtgcaacctggcctttgtctgcatctgcatggaaatc taactgtctcatgtcatcagttgatatccataccaagagccaatatcaacattgcatcattga ctggcaactggggcaggtgtgctgtcgtggtttttcactgtctttgccagtgagctgtcag tctacactcagacgtatcaccttggaaagatggcatcacatcacatcagcatgccatcgctggac tgcgaagtgagctccgcctcgtccagtgctcatgggtgtagtggcctggatgatttggct
FSHui	FSH	T	C	Y	Y	cds	GE611	GGAACCTCCAC AATACCATAC CTA	TTTATCTTTCA TTTACCAACAG G	GGAACCTCCAAATACATAACCTAACTCTCTCTTAACTCCTCAGGATCTGGTGTATAAGGAC ccagccagggcccaaatccagaacaatgctcctctcaagactcagatggtgata [t/c] gaaacagtga gagtcctccggtgtgctcaccatcagatctcctgtgtatataatcccaagtgccacacagtgatc tgtggcaagtgtgacagcgacacgactgtgtactgtgagggcctgagggccagcactgctc CTTTGGTGAATGAAGAATAA

FIG. 5SSS

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
FSHu2	FSH	G	T	S	I	cds	GE561	TTTCTCAGTTT CTAGTGGGTT C	GGTACCTACCC TGGGTAGCA	TTTCTCAGTTTCTAGTGGCTTCattgtttgtcttccagaccaggatgaagacaactccagttttt cttctctttctgtgtcgtgaaagcaaatgtgtgcaata(g/t)ctgtgagctgaccaacatcaacc attgcaatagagaagaaagtgtgtttctgcataagcatcaaccaccttggtgtgtggtgcta CTGTACACCAGGCTAGGTACC
GABRB1 a7	GABRB1	A	G	-	-	noncoding	GE1089	CAGATGGTATT CAAAATGATT CCTA	AACATCTCTT TCCTTTTCACAG	CAGATGGTATTCAAAATGATTTCCTAaacttbtgtttaaccgtgtctgtttatttggtttccagatg tctataccactgatcatgaatttactggaaatgagagagggcgccagtcacctgggttcttaaat aaaatcgaaacttctcaatttcaattgtctgaactagaagatggctgtgtgaaagattcaac aacaggtaggttggttctcccca(a/g)atgtactagggtgtgCTGTGAAGGAAGCAAGATGGTT AACATGTTTGGGCTAGGAAGcctggaaatgaaattgcatcacttttgaattgtttttttttt ctctctctctctatcagatacaacacagctgcattgatcatgttctgtcttcgaagatataccac tggatgagcaaaactcacacctggagat(c/a)gaagattgtgatttacttgcacaggagaaatga aaagagggtttcttcttcttgcaccgttgaaTTCAACTTCTACCTTGAATTATTCACA
GABRB1 a8	GABRB1	C	A	I	I	cds	GE1271	AACATGATTG GGGCTAGGA	TCGTAATTAC TCAGTAGAAG TTGAAT	AACATGATTGAGAGcctgttctaatgttgccccactccac(g/a)lscaggggccc ccgttcagcttggggaagcgatgtcttcagcagatagacatggtcttcgaagatgaattgggtg agtggcctccaggaggccgggttcgggtTACGCAGATGGGAATGGAC
GABRB1 d3	GABRB1	G	A	-	-	noncoding	GE1035	CCAGCTGTG TCACTGAG	GTCCATTTCOC ATCTGCGTA	CCAGCTGTGCTACATGAGaactgttctaatgttgccccactccac(g/a)lscaggggccc ccgttcagcttggggaagcgatgtcttcagcagatagacatggtcttcgaagatgaattgggtg agtggcctccaggaggccgggttcgggtTACGCAGATGGGAATGGAC
GABRB1 d4	GABRB1	C	T	I	I	cds	GE1089	CAGATGGTATT CAAAATGATT CCTA	AACATCTCTT TCCTTTTCACAG	CAGATGGTATTCAAAATGATTTCCTAaacttbtgtttaaccgtgtctttatttgggttccagatg tctataccactgatcatgaatttactggaaatgagagagggcgccagtcacctgggttcttaaat aaaat(c/t)gaacttctcaatttcaattgttgaactagaagatggctgttcgaagatggaggt tcacaacaggtaggttggttctcccca(aat/g)tactagggtgtgCTGTGAAGGAAGCAAGATGGTT
GABRB1 d5	GABRB1	A	G	-	-	noncoding	GE1134	CTCTCAATCT TGAAAAGGA	GAGAGCCCGAG GCATCAT	CTCTCAATCTTGAAAAGGAacttaacagtcggc(a/g)cttcacgtcagtggtgtgtcttctct ttcacaggaaatcacagggtgcttacaatgacaacactcacaccactcaggagagacctccgc aaagatcccttatgtcaaggatgatatttatctgatgggttgctttgtgtgtgtctctggg ctctgtggagatgaccttggtaattacattctcttcttgggaaggccccccagaaaagggagct agcaaacagacagagtgccaatgagaagataaactggagatgaactgaagtcacaggtaagata ttaaattctccacaattctgtttaaatttatcagcatcTGAATGCCCTGGGCTCTC
GABRB1 d6	GABRB1	G	A	L	L	cds	GE1144	TGAAAACAGC AAAGGTC	TGGGACCTGT AAGGTTAAAA	TGAAAACAGCAAGTCTCgaacttbtgttcagagctgttcttttggcatcagggtcgagccc cacggttaacttctctcagccctggaaatcggaatgagcagtgagctgtcggaagtgctcac gaggtgagcgacctcaaggccacctgactcctatgacagccagctcagctcacgcgaag ccctgagcagccgagggcctcagggcgccctggacccgagcgaggtaccacgaaggggcgc atccggagggtgctccacagctaaagtcaagatccccgactt(g/a)actgatgtgaattcca tagacaagtggttcccgaaatttttccccatacccttctcttctttaaigtctgtatctggctt tactatgtcacatgaggtgtgttcaatgggttccatttagactacttctctcttctattgtgttt TAACCTTACAGGTCCTCCA
GABRB1 u1	GABRB1	G	A	T	T	cds	GE1134	CTCTCAATCT TGAAAAGGA	GAGAGCCCGAG GCATCAT	CTCTCAATCTTGAAAAGGAacttaatagtcggcacttcagctaaagtggtgtctctcttcca caggaatcacgac(g/a)tgcttacaatgacaacatcagcaccacctcagggagagacctccgc aaagatcccttatgtcaaggatgatatttatctgatgggttgctttgtgtgtgtctctggg ctctgtctggagtgtgccttggtaattacattcttcttgggaaggccccccagaaaagggagct agcaaacagacagagtgccaatgagaagataaactggagatgaataagtcacaggtaagata ttaaattctccacaattcttctttaaatttatcagcatcATGATGCCCTGGGCTCTC

FIG. 5TTT

FIG. 5WWW

[illegible]

FIG. 5AAAA

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence
GRIN1u 4	GRIN1	G	C	E	Q	cds	GE1287	GAGCTGAGAG AGACTGCCG	GTAGCTGCCGG CGGTAACT	GAGCTGAGAGAGACTGCCGccctggcagccttagtgggtgcacaggctgggtcctcccttc306 ccccccagattgtgagagaccccccag/g/cjagcccttcgtacgtcaagcccccagctgagtgga tggaacatgcaagagagattcaacagcaacggccaccagctcaagaggtgatctctcaccagggc ccaacacagctcgcggcagcctgagtcgcccgggcagggcgccggcgccggcgccggcgccg ggggcggtggggcggtctggagccacagcagttaccgcccgcacacctac
GRIN1u 5	GRIN1	T	G	C	G	cds	GE1287	GAGCTGAGAG AGACTGCCG	GTAGCTGCCGG CGGTAACT	GAGCTGAGAGAGACTGCCGccctggcagccttagtgggtgcacaggctgggtcctcccttc306 ccccccagattgtgagagaccccccag/g/cjagcccttcgtacgtcaagcccccagctgagtgga aca/t/gjcaagggagattcaacagcaacggccaccagctcaagaggtgatctctcaccagggc ccaacacagctcgcggcagcctgagtcgcccgggcagggcgccggcgccggcgccggcgccg ggggcggtggggcggtctggagccacagcagttaccgcccgcacacctac
GRIN1u 6	GRIN1	G	C	D	H	cds	GE1287	GAGCTGAGAG AGACTGCCG	GTAGCTGCCGG CGGTAACT	GAGCTGAGAGAGACTGCCGccctggcagccttagtgggtgcacaggctgggtcctcccttc306 ccccccagattgtgagagaccccccag/g/cjagcccttcgtacgtcaagcccccagctgagtgga acatgcaagggagattcaacagcaacggccaccagctcaagaggtgatctctcaccagggc c/g/cjcaacatcgcggcagcctgagtcgcccgggcagggcgccggcgccggcgccggcgccg ggggcggtggggcggtctggagccacagcagttaccgcccgcacacctac
GRIN1u 7	GRIN1	G	A	K	K	cds	GE1115	GCGGAGCTGG GAGGAC	AGGACGCGAG GTCAGC	GCGGAGCTGGGAGGACGctgctgcacagcccgccctctgcgctcgcaggtgaaacacagc299 aacaagagagattggaatgggagatgagggagagctgctcaacggcgagggcagacatgatctggc ggcgttaacctataaacaagagcgagcgcgccagctacacaggtttcccaagccctcaagtacagg gctgactattctgttcaa/g/a/aaggtggggcgccggcgccggcgccggcgccggcgccggcg gtccctgagggcgccggcgccggcgccggcgccggcgccggcgccggcgccggcgccggcgccg
GRIN1u 8	GRIN1	G	A	E	E	cds	GE1120	TTCCGCGAGTG GGAGGC	CGTCTGTCACC TCGGCT	TTCCGCGAGTGAGGCGgggtggggcgccggcgccggcgccggcgccggcgccggcgccggcg309 cgctcccgagggcagcaccctgagctctctgcacccgctgcgcccctactccaccagtcctgg cgtgtgttgagatgagtgatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg acga/g/a/gggcgccggcgccggcgccggcgccggcgccggcgccggcgccggcgccggcgccg gagggcgccggcgccggcgccggcgccggcgccggcgccggcgccggcgccggcgccggcgccg
GRIN1u 9	GRIN1	G	T	-	-	noncoding	GE1196	ATTCAGGTGG CCAAATTAT	AAAGAAACAA AAACATGTC	ATTCAGGTGGCCAAATTATtggtaagaaactgaaatctaatataataataataataata802 tctaataatttttattatttagtttagtttagtttagtttagtttagtttagtttagtttagtttag gagggaggggcttactgcagctttacatgcaatttataataatattgtaataatattgtaataatattg tgtaaataaagatgatttttagatgagattttttagatgagattttttagatgagattttttagatg g/t/gtcaagaaatgctgagatgagatgagatgagatgagatgagatgagatgagatgagatgag gagcgatggtctcagaacacacagtttgccttaggggagggagggagggagggagggagggaggg tgtgagtgaggttgcagggctgctgagggctgctgagggctgctgagggctgctgagggctgctgag catttgacccctctctatcccaagtgagctgctgagggctgctgagggctgctgagggctgctgag ggcaccatataagcggttactttcacatcacgctcccccagcagttgagttactcaatctaa gcttgagaggttgggaatgatttgcataagaggtaccagatattgtaaatagtcagagatctca taggttgcaataataacatattcttttatacccaacacagaggtttatttccaaataataatg aggaatggtttgtttgtttgtttgtttgtttgtttgtttgtttgtttgtttgtttgtttgtttgttt
GRIN1u 10	GRIN1	G	T	-	-	noncoding	GE628	GACCTCCCAT TACAGTTAT	GAGAAACAC AAAGTTTAT TAGTTTC	GACCTCCCATTACAGTTATtctatgtattt/g/t]tttaataaccacagctcgaaatacaaa383 agaaataataaagatttcacggcgccagggctcagaggtctcacaagaaacccctgaaatcct gttaacaaacacatagttcttcgacgcttaccacacacacacacacacacacacacacacacac ggttattgaaactgaggttataatgagggatattgagagctctgtctccagctcactctgttga tcatgactacgctcaacatgtagggagggcgaggtgaggtgaggtgaggtgaggtgaggtgaggtg ataccaggtgaagatgcaaaacataaagacaaactatataataaactttgtgtgtgtgtgtgtgt

FIG. 5BBB

FIG. 5GGGG

92/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
HSDB2 d28	HSDB2 2	C	G	-	-	noncoding	GE665	CCTGCTGGAAA TAGTGAGCTTC	GCTCTTTTGT TGAACGTGTGTG AA	CCTGCTGGAAATAGTAGCTTCCTactcagcccaatttacaccccttcaacccgccc acacagtcacattatcaaatagcgtattccctctctcaagaagctcagcagatctggcg tataagccactctacagctggagagcccaagcagagaaacggaggagtggttgccttgt ggacggcacagagagacccctgaagctccaaagctcagtgatttaaggatcacagatgcatg tgggtattgttagaaatgcatcaaatccaccctggccctcacagaaagcaacaaagggc acaagccaggctcctgctccctttcaccaatgcccaacttactgtattccctcagtcac aaaacccgacagtcacggcccaacaaagagcgtttctgctcctaactatcacccagagaaaac aatagtattgtgttaccaaatctcagtagctgattctgaacattgagggacccttaaatg aagggcccttttgactaatagagctccattccact(c/g)ttaatgagaagcatttcctttc tctttaatctcccatctcctTCACACAGTTCAACAAAAGAGC
HSDB2 d29	HSDB2 2	C	T	-	-	noncoding	GE665	CCTGCTGGAAA TAGTGAGCTTC	GCTCTTTTGT TGAACGTGTGTG AA	CCTGCTGGAAATAGTAGCTTCCTactcagcccaatttacaccccttcaacccgccc acacagtcacattatcaaatagcgtattccctctctcaagaagctcagcagatctggcg tataagccactctacagctggagagcccaagcagagaaacggaggagtggttgccttgt ggacggcacagagagacccctgaagctccaaagctcagtgatttaaggatcacagatgcatg tgggtattgttagaaatgcatcaaatccaccctggccctcacagaaagcaacaaagggc acaagccaggctcctgctccctttcaccaatgcccaacttactgtattccctcagtcac aaaacccgacagtcacggcccaacaaagagcgtttctgctcctaactatcacccagagaaaac aatagtattgtgttaccaaatctcagtagctgattctgaacattgagggacc(c/t)ttaa actgagggcccttttgactaatagagctccatttcctctcttaaatgagaagcatttcctttc tctttaatctcccatctcctTCACACAGTTCAACAAAAGAGC
HSDB2 u1	HSDB2 2	C	G	R	R	cds	GE639	AAATATAGGCA TCTGCTGAGTG TAT	CCATCAGAGT TTAAGATGGAG	AAATATAGGCAATCTGCTGAGTGTTAaccatttaccctctgttttagccctctctgggtcac gctgaatcagatctgctctccagcactctctgttctcctggcaagtggttccctgctactttgga tggccagatgacggctggagctgctctgttgacagagagaggggcttctgggtcagggatc gtccg(c/g)ctgttggtggaagaaaggaactgaagagatccagggccttggaacaggcctca gcccagaattgagagaggaattttctgtagtaagtaaaactgagtcagtcgtgtgtgCTCATT AAACTCTGCATGG
HSDB2 u10	HSDB2 2	A	G	-	-	noncoding	GE665	CCTGCTGGAAA TAGTGAGCTTC	GCTCTTTTGT TGAACGTGTGTG AA	CCTGCTGGAAATAGTAGCTTCCTactcagcccaatttacaccccttcaacccgccc acacagtcacattatcaaatagcgtattccctctctcaagaagctcagcagatctggcg tataagccactctacagctggagagcccaagcagagaaacggaggagtggttgccttgt ggacggcacagagagacccctgaagctccaaagctcagtgatttaaggatcacagatgcatg tgggtattgttagaaatgcatcaaatccaccctggccctcacagaaagcaacaaagggc acaagccaggctcctgctccctttcaccaatgcccaacttactgtattccctcagtcac aaaacccgacagtcacggcccaacaaagagcgtttctgctcctaactatcacccagagaaaac aatagtattgtgttaccaaatctcagtagctgattctgaacattgagggacccttaaatg actgagggcccttttgactaatagagctccatttcctctcttaaatgagaagcatttcctttc tctttaatctcccatctcctTCACACAGTTCAACAAAAGAGC

FIG. 5III

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence	
HSD3B2 u15	HSD3B 2	C	T	-	-	noncoding	GB665	CCTGCTGGAAA TAGTGAGCTTC	GCTCTTTTGT TGAACGTGTG AA	CCTGCTGGAAA TAGTGAGCTTC Cctaccagcccaattacacctatacaaccccccttcaacgcgc acacagccacattatacaaatagcgtatccctctctctacagaagcctcagcgagatctggcg tataagccactctacagctggaggaagcaagaaacccgtggagtggtggttcccttgt ggaccggcaagagagaccctgaagtcacagctcagtgattgaagatgacagagatgcatg tgggtattcttaggaatgtcatcaactccaccctggcctacacagaagcaacaagggc acaagccaggctccgctgctcccttcaacaatggcccaacttactgtattccctcagtcac aaa(c/t)ctgcacagtcactggcccaagaagcttctgctcctcaatcacaccagaagaa aaacaatgatttggcttaccacactcagtagctgattctgaacaattggggacccttaa actgaagggcccttttgactaatagagctccattccactcttaaatgagaagcattctcttc tcttaattctccattcttCACACAGTTCACAAAAGAGC	627
HSD3B2 u16	HSD3B 2	A	C	-	-	noncoding	GB665	CCTGCTGGAAA TAGTGAGCTTC	GCTCTTTTGT TGAACGTGTG AA	CCTGCTGGAAA TAGTGAGCTTC Cctaccagcccaattacacctatacaaccccccttcaacgcgc acacagccacattatacaaatagcgtatccctctctctacagaagcctcagcgagatctggcg tataagccactctacagctggaggaagcaagaaacccgtggagtggtggttcccttgt ggaccggcaagagagaccctgaagtcacagctcagtgattgaagatgacagagatgcatg tgggtattcttaggaatgtcatcaactccaccctggcctacacagaagcaacaagggc acaagccaggctccgctgctcccttcaacaatggcccaacttactgtattccctcagtcac aaa(c/t)ctgcacagtcactggcccaagaagcttctgctcctcaatcacaccagaagaa aaacaatgatttggcttaccacactcagtagctgattctgaacaattggggacccttaa actgaagggcccttttgactaatagagctccattccactcttaaatgagaagcattctcttc tcttaattctccattcttCACACAGTTCACAAAAGAGC	627
HSD3B2 u17	HSD3B 2	G	T	L	L	cds	GE1194	CAGAAGAATGC ACCTTGAGTC	GCCAGATCTCG CTGAGCC	CAGAAGAATGC ACCTTGAGTC Tctaaacacccacacgagggaggggacacacga gcacagcagtgagcagcactccgggagaaatccctcaacacacacacacacacacacac taccagctacttg/l tggggcccttggttccagccagtgccagctctctcatcaccagtg agctatagagtgagccggcccaactctccacaggaatcatccagaaagccacaggaagagcc tctggaacacacatggcccaactccacccgacagcaaaaagcttggtagaaggtgctggg cggtaatgggtggaatctaaaaatgggtgataccttgacacttggcttgaagccacatat atctatggggaagggagcccatctctctgcccagtaataatgagccctgacacacaaatgggat cctgcaatggttgaaaagtctccacagtcacccagctctatgttggcaagctggccctgggccc acattctggccctggggccctcgaggcccccagaagggcccccagctcagggacaaatctac tacaatctcagatgacacgctccacaaagctatgaaccccttaattacacctgagcaagagtt cggtccgcttgattccagatggagccttcttaacctgtatgactgattggcttctctgc tggaaatagtgagcttccctcagcccaattacacctatacaaccccccttcaacggccacaca gtcacattatacaaatagcgtatccacttctctcaagaagggctCACGAGATCTGGC	839
HSD3B2 u18	HSD3B 2	C	T	L	L	cds	GB665	CCTGCTGGAAA TAGTGAGCTTC	GCTCTTTTGT TGAACGTGTG AA	CCTGCTGGAAA TAGTGAGCTTC Cctaccagcccaattacacctatacaaccccccttcaacgcgc acacagccacattatacaaatagcgtatccctctctctacagaagcctcagcgagatc/l ggcgtataagccactctacagctggggaagcaagcaagcaaacccgtggagtggttcccttccc ttggggccggcacaagagagaccctgaagtcacagctcagtgattgaagatgacagagatg catgtgggtattttaggaatgtcatcaactccaccctggcctacacagaagcaaca gggcaaacgcccaggtcctgctgctcccttcaacaatggcccaacttactgtattccctcagtcac catcaaacctcagcagtcactggcccaagaagcttctgctcctcaatcacaccagaagaa aaacaatgatttggcttaccacactcagtagctgattctgaacaattggggacccttaa actgaagggcccttttgactaatagagctccattccactcttaaatgagaagcattctcttc tcttaattctccattcttCACACAGTTCACAAAAGAGC	627

FIG. 5KKKK

SUBSTITUTE SHEET (RULE 26)

FIG. 5TTT

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/ noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence	
HTR1D6		A	G	T	T	cds	GE1162	TTGAAGGAGG AGCCAAATG	TGTAGGAGATC TGAGAGGTGTT CA	TTGAAGGAGGAGCCAAATGtgtggaggtctgtgggagagagagagccactgcagcatgtcccact gaaccagtcagcagaaggcctcccagaggcctcccagagctccacagatccctgaatggccagaaaact cagaggtctgggataccagagacctccagggcctcaagatccctctcgctggctctcttcgcgc atcacl(a/g)ctggccacagtcctctcaatgctcttgtaactcaacacattactacacagga agctccacacctccgcacactcattggtctccctggccaccacgacacctctgggttccatc tctggttaagcccatcagcatcctctatccatccacacaccttggaacattggccaaaattttgtg tgacatcggtctctctcgacatcagctgtgcacagctccctctctctctctgtgcatatg ctctgcagagtaactggcaatcacagatgccctgggaatagagtaaacagcagagcgtgtggcac gcggccaccatgatgcccatgtctgggccaatctcatctgcatactccaccgcgcgtctcttg gcggcgccgaagccccagagagagatgtcggagctgtctgtgTGACACCTCTCAGATCTCCTACA	650
HTR1D6		T	C	I	I	cds	GE1162	TTGAAGGAGG AGCCAAATG	TGTAGGAGATC TGAGAGGTGTT CA	TTGAAGGAGGAGCCAAATGtgtggaggtctgtgggagagagagcactgcagcatgtcccact gaaccagtcagcagaaggcctcccagaggcctcccagagctccacagatccctgaatggccagaaaact cagaggtctgggataccagagacctccagggcctcaagatccctctcgctggctctcttcgcgc atcaactggccacagtcctctcaatgctcttgtaactcaacacattactacacaggaagct ccacacacctgccaaataactgat(t/c)ggctccctggccaccacgacacctctgggttccatc tctggttaagcccatcagcatcctctatccatccacacaccttggaacattggccaaaattttgtg tgacatcggtctctctcgacatcagctgtgcacagctccctctctctctctgtgcatatg ctctgcagagtaactggcaatcacagatgccctgggaatagagtaaacagcagagcgtgtggcac gcggccaccatgatgcccatgtctgggccaatctcatctgcatactccaccgcgcgtctcttg gcggcgccgaagccccagagagagatgtcggagctgtctgtgTGAAACACCTCTCAGATCTCCTACA	650
HTR1D7		C	G	S	S	cds	GE1162	TTGAAGGAGG AGCCAAATG	TGTAGGAGATC TGAGAGGTGTT CA	TTGAAGGAGGAGCCAAATGtgtggaggtctgtgggagagagagcactgcagcatgtcccact gaaccagtcagcagaaggcctcccagaggcctcccagagctccacagatccctgaatggccagaaaact cagaggtctgggataccagagacctccagggcctcaagatccctctcgctggctctcttcgcgc atcaactggccacagtcctctcaatgctcttgtaactcaacacattactacacaggaagct ccacacacctgccaaataactgatgttggctccctggccaccacgacacctcttgggttccatctgg taatgcccatacagcatcgtctataccatcaccaacactggaaacttggccaaaattttgtgtagac atctggctgt(c/g)tcctgacatacagctgtgcacagctccctctctctctctgtgcatatg ctctgcagagtaactggcaatcacagatgccctgggaatagagtaaacagcagagcgtgtggcac gcggccaccatgatgcccatgtctgggccaatctcatctgcatactccaccgcgcgtctcttg gcggcgccgaagccccagagagagatgtcggagctgtctgtgTGAAACACCTCTCAGATCTCCTACA	650
HTR1D8		A	C	A	A	cds	GE1162	TTGAAGGAGG AGCCAAATG	TGTAGGAGATC TGAGAGGTGTT CA	TTGAAGGAGGAGCCAAATGtgtggaggtctgtgggagagagagcactgcagcatgtcccact gaaccagtcagcagaaggcctcccagaggcctcccagagctccacagatccctgaatggccagaaaact cagaggtctgggataccagagacctccagggcctcaagatccctctcgctggctctcttcgcgc atcaactggccacagtcctctcaatgctcttgtaactcaacacattactacacaggaagct ccacacacctgccaaataactgatgttggctccctggccaccacgacacctcttgggttccatctgg taatgcccatacagcatcgtctataccatcaccaacactggaaacttggccaaaattttgtgtagac atctggctgtctgacatacagctgtgcacagctccctctctctctctgtgcatatgctct ggcaggtactctggg(c/a/c)atcacagatgccctgggaatagagtaaacagcagagcgtgtggcac gcggccaccatgatgcccatgtctgggccaatctcatctgcatactccaccgcgcgtctcttg gcggcgccgaagccccagagagagatgtcggagctgtctgtgTGAAACACCTCTCAGATCTCCTACA	650

FIG. 5UUU

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/ noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence	
HTRI1D1	HTRI1D	G	T	R	R	cds	GE1162	TTGAAGGAGGAG AOCCAAATG	TGTAGGAGATC TGAGAGGTGT CA	TTGAAGGAGGAGCCAAATGtctggaggtctctgggagagagagccacctagcatgtcccactgacacgctcagcagaagcctcccagagggctcccaacgagtcccttgatggccacagagaacctcagagcctgggattggagccctccagggcctcagatctcccttgcctgggtctgttcgcgtctccgctatcacacggccacactctccaaagtcttgactcacaccactcttactaccagagaagctatcacacccctgcacaactactgattggctctctggccaccaccgacctctgggttccactctgggtaatgacatcagatcctaccacacacctggcaacctgtggcacaactcttggtgacatctggctgtctctgacatcgtgtgcacagctccatctgcatctgtgattgtctgtctggagactggcgaatcacagatgctctggaataacagtaaacagagacagctgtgtggccaggggacacatgatcgccatgtctgggccaactccactgtctccatccatccccctctctggcgctg/tcaggccaaggccccaggagagatgtcggaactgtctgtgTAACACCTCTCAGATCTCTCTACA	650
HTRI1E1	HTRI1E	C	T	I	I	cds	GE1160	TTCCCTTGTTA CAGGTATCCAT	GTGGAGGTAG AAAGCTCCA	TTCCCTTGTTACAGGTATCCATtttcagctataatactttttaaacaagaagaatggattctttaaattcatctgatcaaaacttgacctcagaggaactgttaaacagagatgccatccaaatcttgggtctccctactctgtctgggtcggcactgatgacaacactcaactccctctggtgat (c/t)gctgcaattattgtgaccgggaagctgcaccatccagccaattatttaatttgcttccctggcagctcacagattttctgtggctgtctcgtggatgcctcagcaattgtatattgttgagagagagctggattatggggccaagggtctgtgacatttggctgagtggtgacataactgctgcacggctcccattctgtgactctcagctatgcttgggtggatcagcaatccacagactcagatgctgtggatgtgcaggaaaaggactcccaagcagctgtggcattatgatacaatagtttggaatactattgttttatctctatgctctctattctgtggagcaccaggaactcagagatgatgaatgcatacaagcacyaccacattgltttccaccatttactcaacattTGGAGCTTCTACATCCAC	640
HTRI1E2	HTRI1E	T	C	I	T	cds	GE1160	TTCCCTTGTTA CAGGTATCCAT	GTGGAGGTAG AAAGCTCCA	TTCCCTTGTTACAGGTATCCATtttcagctataatactttttaaacaagaagaatggattctttaaattcatctgatcaaaacttgacctcagaggaactgttaaacagagatgccatccaaatcttgggtctccctactctgtctgggtcggcactgatgacaacactcaactccctctggtgatcgtgcaatta (t/c)gtgaccgggaagctgcaccatccagccaattatttaatttgcttccctggcagctcacagattttctgtggctgtctcgtggatgccttcagcaattgtatattgttgagagagagctggattatggggccaagggtctgtgacatttggctgagtggtgacataactgctgcacggctcccattctgtgactctcagctatgcttgggtggatcagcaatccacagactcagatgctgtggatgtgcaggaaaaggactcccaagcagctgtggcattatgatacaatagtttggaatactattgttttatctctatgctctctattctgtggagcaccaggaactcagagatgatgaatgcatacaagcacyaccacattgltttccaccatttactcaacattTGGAGCTTCTACATCCAC	640
HTRI1E3	HTRI1E	T	G	V	G	cds	GE1160	TTCCCTTGTTA CAGGTATCCAT	GTGGAGGTAG AAAGCTCCA	TTCCCTTGTTACAGGTATCCATtttcagctataatactttttaaacaagaagaatggattctttaaattcatctgatcaaaacttgacctcagaggaactgttaaacagagatgccatccaaatcttgggtctccctactctgtctgggtcggcactgatgacaacactcaactccctctggtgatcgtgcaatta (t/c)gtgaccgggaagctgcaccatccagccaattatttaatttgcttccctggcagctcacagattttctgtggctgtctcgtggatgccttcagcaattgtatattgttgagagagagctggattatggggccaagggtctgtgacatttggctgagtggtgacataactgctgcacggctcccattctgtgactctcagctatgcttgggtggatcagcaatccacagactcagatgctgtggatgtgcaggaaaaggactcccaagcagctgtggcattatgatacaatagtttggaatactattgttttatctctatgctctctattctgtggagcaccaggaactcagagatgatgaatgcatacaagcacyaccacattgltttccaccatttactcaacattTGGAGCTTCTACATCCAC	640

FIG. 5VVV

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/ noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
HTRIEL L	HTRIE L	A	T	E	V	cds	GE1160	TTCCCTTGTTA CAGGTATCCAT	GTGGGATGTAG AAGCTCCA	TTCCCTTGTTACAGGTATCCATLTTCCAGCTATATTAACTCTTTTAAACCAAGAGAAATGGATTCTTAAATTCATCTGATCAAAACTTGACCTCAGAGGAGTGTAAACAGAGATGCCATCCAAAATCTTGGTCTCCCTCACTGTCTGGCTGAGTCAACCAACTATCACTCCCTTGTGATCGTGCATATTGTGACCCGGAAGTCCCACTCCAGCAAAATTAAATTAATTTCTCCCTTGCAGTCAACA GATTCTTGTGTCTCTGGTGCCTTCAGCATTTGTATATTGTGAGAGAGAGCTGGATATGGGCAAGTGGTCTGTGACATTTGGCTGAGTGTGACATTTACCTGTGCACTGCTCCCATCTTGCATCTCAGTATAGCTTTGGATCGGTATCGAGAAATCAACAGATGCGTCTGTGATGTCAGGAAAGGACTCCCAAGCATCTGGCTATATGATTAACAATGTTGGATATATATCTCTTTTATCTCTGCTCTCTATCTGTGGGCAACCAAGCACTAGCAGAGATGAGTGAATGATCATCAAG CACGACCACTGTCTCCACCATTTACTCAACATTTGAGCTTTTCTACATCCCAAC
HTRIEL L	HTRIE L	T	A	I	I	cds	GE1158	CGACCACTTG TTTCACCA	AGTTATCTCTC CCCTCAAAA	CGACCTTACTCAAAATATAGTCAACCAATTTGGAGCTCTTATCCCAAGACAGCAAGTAGG TGAATCTCATCATCAAAACTTGACCTCAGGGAAGTGTGAGAGAGTGTGAGAACTCAAACTCAGT TTCCATCTCATGTACTAGAAAGCTTTTATCTGACCATCAACAGACTTTGATAAAAT [t/a] CATAGCAGTGGAAAGTCTCAGGTCGAAATCAAGCATGAGAAATCTGGGAAGGCCAAAGAT CTGCGTCAAGGAAAGGAGCGCCACTACCTGGGTAACTAATCTTGGTGCATTTGTCAATAT GTGGTCTCTCTTTTGTAAAGAAATAGTGTAACTGTGTGCAAAATGAAATCTCTGGA GAAATTCGAATTTTGGCATGGCTGGTGTATCTATCTCTTAAATCACTCACTGAGTCTACAC AATCTTAAAGAGACTTCAAGAAAGCATCCAAAAGCTTGGCATGTGAGTGTGATTTTAAa AATGCTTATCTATGGAAGTGGGGTTTTCAGGGGAGGAATAACT
HTRIEL L	HTRIE L	T	G	D	E	cds	GE1160	TTCCCTTGTTA CAGGTATCCAT	GTGGGATGTAG AAGCTCCA	TTCCCTTGTTACAGGTATCCATTTTCCAGCTATATTAACTCTTAAACCAAGAAATGGATTCTTAAATTCATCTGATCAAAACTTGACCTCAGGGAAGTGTGAGAGAGTGTGAGAACTCAAAATCTTGGTCTCCCTCACTGTCTGGCTGAGTCAACCAACTATCACTCCCTTGTGATCGTGCATATTGTGACCCGGAAGTCCCACTCCAGCAAAATTAAATTTCTCCCTTGCAGTCAACA GATTCTTGTGTCTCTGGTGCCTTCAGCATTTGTATATTGTGAGAGAGAGCTGGATATGGGCAAGTGGTCTGTGACATTTGGCTGAGTGTGACATTTACCTGTGCACTGCTCCCATCT TGCATCTCAGCTATAGCTTTGGATCGGTATCGAGCAATCAACAGATGCGTGTGATGTGCCAGG AAGAGGACTCCAAAGCATCTGGCATATGATTAACAAGTGGATATATATGTTTATCTCTCTGTGCTCTCTATCTGTGGGCAACCAAGCACTCAGAGAGATGATG/GAGATGATCATCAAG CACGACCACTGTCTCCACCATTTACTCAACATTTGAGCTTTTCTACATCCCAAC
HTRIEL L	HTRIE L	T	C	I	T	cds	GE1157	CAGCCAAAGGA AAATAACCA	GCACCCAGCTT GGAGTAAAT	CAGCCAAAGGAATAACCAACAGCTTCTCCAGCTGTAGACTGAACAGAGGAACATGAACAT CCAAACTGTACCAAGGAGGCAAGTGGCTATGACCAAGCAAGCACTCACTGTGAGATGCTCA TTTGTGACTCTGGTGGTCTACCAACCTCCCAAGTGTGCTGAATTTGGCTGAAGATGCTCA TATTGGCATTTGGCAACCAAGAAAGCTCCACAGCTCCCAACTACCTATCTGTCTCTGGCGCTGA CGGACCTCTGGTGGCAGTCTGCTCATGCCCTGAGCATCATCATCTGTATGATGTGATGCTGTGA AAGTTGGTAACTCTCTCTGTGAGTGTGGGTAGTGTGGCACTGACACTGCTCACTGCTCCAT CCTCCACTCTGTGCTATGGCTGGCAGGTAATGGGCAATCAACCAATCTATTGAATACGCA GGAAGAGGAGGCGCAAGGCGCGCTGATGCTTACGCTTGACCATCTCCATCTTCTCAT TCCATGCCCTCTGTCTGGAGAGCCACCGCGCTAAGCTCCCTCCCTCCCTAGCTAGTGCACAT CCACGACCACTGTTATCTACACTTTACTCCACGCTGGGCTGC

FIG. 5 WWW

FIG. 5YYYY

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
HTR2Ad 14	HTR2A	C	T	-	-	noncoding	GE1167	TGGTACTGCGA AACCA	CGCACTGCTAG GATCCTGT	TGGTACTGCGAACCACCACTatttctaccacatgtgaggttttgaataataatctgggtggcata ttctctggaagaataagccagttcaatgtgtatctctataataataaataagtgctaatagtt tatcagagttatcacccacagactgctcagccctcagcctatgtggccaatgtcagtaattcc acttggacacacacactgttggctttggatggagtgccagacactccagcctc/c/19ag gacatactgtttctcagctctatttcttctcctccctcagctcctcaaaaaattaccag tgccacttactacttattgggaatctgcacaaaggccatttctccagtttctctcaaaagca aagaaaaattcccaataataatagcaaaaggaggaagaaagccgttttggccgc cctcctggtgtgtctacactgttattgacatacacatagagggaggtctgtatgaatgaac ggacagtcagagagctactcctcctcctggaacacagagtcctctggtgcagacactcttc ctacttcccatgagttctttgtgcgactttgaggggtctgtgaatgatttctaaatgtgc ctactgagcgagccgcagagggaggaacccagccgctgcagaggaagccAACAGG ATCCTAGCAGTCGG
HTR2Ad 15	HTR2A	G	A	-	-	noncoding	GE1169	CCGTGCCAGAG GAAGCC	CTCACCACAC GAGGACAAA	CCGTGCCAGAGAACCACTatttctaccacatgtgaggttttgaataataatctgggtggcata ttgaagtcagcaaaacagaaacacaaattactatctatatttctgtgtggaagatcaagagaggg gactctacacccgttttaattctgtgagagtcagcagcagcagaaatcaaaatgtatccat gtgtgaacccctggaagacaaatgtatgtctcctcagcctataatttattgtgtgtaatttctt tccggtttgaaatcatctgtggccaaatgtaacttcaatgagaatttccagggaggaaggt tgtctgtaacttttacttaagactttttgttttcttcttatttagcaagacattataggag ctgaaattcctgacagcgtgtgcaattcagcctgaatggctgagaactgaacccaaga tacaatcaattactatgggttaacactggaatgtatttttaattgacttcttaattgagaatgt gtaatcccaactgttttctgtatgtctgtatttataataactgttcttaactgaccactt g/a/gcataacacaaatgagatagtttaacagaggtccagagttataaaactttctt cttggcagacatttcttcccgagcgtcaaaaaaaccttgaacctctatgtctataaag tttcaattctgtcttttTGTGCTGGTGTGGAG
HTR2Au 1	HTR2A	G	C	-	-	noncoding	GE1167	TGGTACTGCGA AACCA	CGCACTGCTAG GATCCTGT	TGGTACTGCGAACCACCACTatttctaccacatgtgaggttttgaataataatctgggtggcata ttctctggaagaataagccagttcaatgtgtatctataataaataagtgctaatagtt tatcagagttatcacccacag/c/c/actcctagccacccctcagcctatgtggccaatgtcagtaa ttccactctggacacacacactgttggctttggatggagtcgacacactcagcactccgag gacatactgtttctcagccttatttcttcttctccttccactcagcctcaaaaaattaccag tgccacttactacttcaatgggaatctgcacaaaggccatttctccagtttctctcaaaagca aagaaaaattcccaataataatagcaaaaggaggaagaaagccgttttggccgc cctcctggtgtgtctcactgtcagcttattgacatacacatagaggggtgtgtatgaatgaac ggacagtcagagagctactcctcctcctggaacacagaggtcccttgggtgcagagctcttc ctacttcccatgagttctttgtgcgactttgaggggtctgtgaatgatttctaaatgtgc ctgtgagggcgagccgcagagggaggaacccagccgagcctgccaagaggaagccAACAGG ATCCTAGCAGTCGG

FIG. 5ZZZZ

[illegible]

FIG. 5A AAAA

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/ noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
HTR2AU 7	HTR2A	A	G	-	-	noncoding	GE1167	TGGTACTGCGA AACCA	CGCACTGCTAG GATCTGTT	TGGTACTGCGAAACCAACTattatttctaccacatgtggagtttttgaataataatctgggtggcata ttttctggaagaaataaagccagttcaatgggtgatctattataataaaaaagtcgtcaataagtt tatcagaattatcacacagactgcttagccacctgagctatggcccaatgtgcagtaattccc actctggacacaaactgttggtttggatggagatgccagactctccgagggaca taactgtttcagcttattatgtctctctccaccttcagcttaaaaaatataccacagtccc atactaccttaatgggaatcgcaaaaaggccattctccagttctctcaaaagcaagg aaaaatccclla/g/ataataatagcaaaagagggagaaagagctgttttggtccgc ccctctgctgttgctacctatgacatacacatagagggaggtctgttgaatgaac gagacagtacagagactcctatccctgggaaccacagaggtctcttggtagacagctcttc ctactttcccatcgactcttttggagactttggagggctcgtaatcttctcaaatgtgc ctgctgagggcgcgcacagggaggagaccacagcgagcgtgcagagggagcchAACAGG ATCTGACGAGTGG
HTR2AU 8	HTR2A	A	G	-	-	noncoding	GE1167	TGGTACTGCGA AACCA	CGCACTGCTAG GATCTGTT	TGGTACTGCGAAACCAACTatttctaccacatgtggagtttttgaataataatctgggtggcata ttttctggaagaaataaagccagttcaatgggtgatctattataataaaaaagtcgtcaataagtt tatcagaattatcacacagactgcttagccacctgagctatggcccaatgtgcagtaattccc actctggacacaaactgttggtttggatggagatgccagactctccgagggaca taactgtttcagcttattatgtctctctccaccttcagcttaaaaaatataccacagtccc atactaccttaatgggaatcgcaaaaaggccattctccagttctctcaaaagcaagg aaaaatcccaataataatgtagcaaaagggagaaagagctgttttggtccgcctc ctggtgtgtctaccttgagcttatgacatacac/la/g/taagagggaggtctgaatgaatgaac gagacagtacagagactcctcctccctgggaaccacagaggtctcttggtagacagctcttc ctactttcccatcgactcttttggagactttggagggctcgtaatgattttcaaatgtgc ctgctgagggcgcgcacagggaggagaccacagcgagcgtgcagagggagcchAACAGG ATCTGACGAGTGG
HTR2AU 9	HTR2A	G	A	-	-	noncoding	GE1169	CGTGCCAGAG GAAGCC	CTCACCAACC GAGGACAA	CGTGCCAGAGGAGGCGCaacaggaatcctagcagtcggaactggtcagctcttcgactgagttt ttgaagtcgcaaaacagcaaaactactactatcatatgtctggtggagaaatcaagaagaggg gacttaacacaggttaatactatggtggagatgcagcgcagtcacagatcaaaatgactctcat gtggaacctggagacaaatgtaagtgttcagtcgcgtatattttatgctgtgtaatttttt tcctgttgaaatcagctgtggcaaatgtaataatcgaatgtaataatccagggaggaagt tgcctgtaactttttactaaagctttttgtttctttttattagctaaagcaaatatagag ctgaatactctgcagcagctgtggcaatca/g/a/ctaagaatggctggagactgttaaccga aagataatccaatctactatggaatcaacactggaatgatttttaattgaactcttcaatgtaga atgtgtacatccccactgttttctgactgcatgtattttaataactctgttcaaatctgtacc atcggtcaacacaaatgagatagttaaacagagccccagctgatttaaaactttttt ctttgtccagacatttatctctcccgacgctcaaaaaaacccctgcacactctatgctaaaag tttcatctgcttttttTGCTCGGTTGGTAG
HTR2CU 11	HTR2C	G	C	C	S	cds	GE1126	TTTTTCAGTGT GCACCTAAATG	ACTTACCNRA AGGATTGCC	TTTTTCAGTGTGCACTTAATTTGgctatttggtttggcgaat/g/ctgatactttctgagcccaag tcagcgtatgtaactcagcaattctcaactcccgagggcgagctccaattcccccagaggg gtcaaaaactggcagcaacttcaactcgtctatcaataaataacatcagcaataggtgtgcaactcct tgtatcatctggcagtagctggaagaaactgcacaaatgcaccaatatttaattgactgtccc tagccat-gctgatatgctagtgaggactacttgtcatgccctctgtctctctctggcCAATCTCTTAT GSTAGT

FIG. 5DDDDD

[illegible]

FIG. 5GGGG

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
IGF1u1 4	IGF1	T	G	-	-	noncoding	GE688	TCATAGCCTAG AAAATGATCCC TAT	AGGGTTTGCAAT CAATTGTGTT	TCATAGCCTAGAAAATGATCCTATCTGAGATCAAGATTTCTCATAGAACATGAATTATCC AGCATTCAGATCTTCTAGTCACTTGAAGCTTTTGTGTTAAAGTACCCAGGCTGATTTCTC ATGCAATCTATATTCTACTCTTGGAGCTCTATATGAAACCAAAATACATCTTCAGTTC TTCTCCCTGGTCCCTCAAGGATCAGAGCCAGGAAAAAAGAAAGACTCCCTGGATCTC TGAATATGCAAAAGAGGCCCATTTAGTGGAGCAGCAATCTGTTCAACACAGTATT TTAACTCTAGTCCCAATATTGTAATGGACCTCCAGCAATG/IGTGGCAATGTTCTAA TCACTAGGACAGATGAAAGAAATATACATCATTTTGGCTCTGCTGTTTCCAGACATA TCACTATGGAATAGATACAGTACCTCTCTCCCAAGATGGCACTCTTTTATTTCTGTC CCAGTGTACCTTTTAAATATATCCCTCTCAACAAATTTATAGGAGTCTATGACTTATCTCTCAG TAACTCTATCCAGTCAAAATCCCAAGGAGGAAAGTGAAGATGCAACTGCCAATATTA CTTTCTTAATCTTCTCAACATATCTCTCACTGATTAATAATGAATGAAATAACT CATTATCACTTCACTATTCTTTTAAATGAATTAAGTAACTAGAAAACAAATTTGATGCAAACT CT
IGF1u1 5	IGF1	T	G	-	-	noncoding	GE688	TCATAGCCTAG AAAATGATCCC TAT	AGGGTTTGCAAT CAATTGTGTT	TCATAGCCTAGAAAATGATCCTATCTGAGATCAAGATTTCTCATAGAACATGAATTATCC AGCATTCAGATCTTCTAGTCACTTGAAGCTTTTGTGTTAAAGTACCCAGGCTGATTTCTC ATGCAATCTATATTCTACTCTTGGAGCTCTATATGAAACCAAAATACATCTTCAGTTC TTCTCCCTGGTCCCTCAAGGATCAGAGCCAGGAAAAAAGAAAGACTCCCTGGATCTC TGAATATGCAAAAGAGGCCCATTTAGTGGAGCAGCAATCTGTTCAACACAGTATT TTAACTCTAGTCCCAATATTGTAATGGACCTCCAGCAATG/IGTGGCAATGTTCTAA TCACTAGGACAGATGAAAGAAATATACATCATTTTGGCTCTGCTGTTTCCAGACATA TCACTATGGAATAGATACAGTACCTCTCTCCCAAGATGGCACTCTTTTATTTCTGTC CCAGTGTACCTTTTAAATATATCCCTCTCAACAAATTTATAGGAGTCTATGACTTATCTCTCAG TAACTCTATCCAGTCAAAATCCCAAGGAGGAAAGTGAAGATGCAACTGCCAATATTA CTTTCTTAATCTTCTCAACATATCTCTCACTGATTAATAATGAATGAAATAACT CATTATCACTTCACTATTCTTTTAAATGAATTAAGTAACTAGAAAACAAATTTGATGCAAACT CT
IGF1u1 6	IGF1	T	C	-	-	noncoding	GE688	TCATAGCCTAG AAAATGATCCC TAT	AGGGTTTGCAAT CAATTGTGTT	TCATAGCCTAGAAAATGATCCTATCTGAGATCAAGATTTCTCATAGAACATGAATTATCC AGCATTCAGATCTTCTAGTCACTTGAAGCTTTTGTGTTAAAGTACCCAGGCTGATTTCTC ATGCAATCTATATTCTACTCTTGGAGCTCTATATGAAACCAAAATACATCTTCAGTTC TTCTCCCTGGTCCCTCAAGGATCAGAGCCAGGAAAAAAGAAAGACTCCCTGGATCTC TGAATATGCAAAAGAGGCCCATTTAGTGGAGCAGCAATCTGTTCAACACAGTATT TTAACTCTAGTCCCAATATTGTAATGGACCTCCAGCAATG/IGTGGCAATGTTCTAA TCACTAGGACAGATGAAAGAAATATACATCATTTTGGCTCTGCTGTTTCCAGACATAAG G/IGTGGTACCTTTTAAATATATCCCTCTCAACAAATTTATAGGAGTCTATGACTTATCTCTCAG TAACTCTATCCAGTCAAAATCCCAAGGAGGAAAGTGAAGATGCAACTGCCAATATTA CTTTCTTAATCTTCTCAACATATCTCTCACTGATTAATAATGAATGAAATAACT CATTATCACTTCACTATTCTTTTAAATGAATTAAGTAACTAGAAAACAAATTTGATGCAAACT CT

FIG. 5JJJJJ

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence	
IGF1u1 7	IGF1	A	T	-	-	noncoding	GE1192	GGCTTAATGAA ATAGCATTAGG T	GATGCCATTCC ATAAATCAGA	GGCTTAATGAAATGACATTAGGTCTctatagccaccaccacatttcaactttttatcactcacaa gtagtctactgtctcccaaatgtgggttcgagggcgaggttggaaatttttttaa gttagaaggtccactgtttttgttgcctcctcaacttagcaaatagcaatattatattccaatc ttctgaacttgatcaagagcatggagaataaacgcggggaataatcttttaggcaaatagaag aatttaaaagataagaaagtctcttctgtttgttgcctctgctctcaaaacagatattcagc aagtgggaataaagaacaaagaataacatacatagattacctgcgaataatagctctcgc aaantcccnnggggaatcttggcaattactgggtttatagagagactctcccttcccccag acattccaaagagcagtagctctcctgaagaaacatcactgactctatttgggaattgtgaaa gtattccctatgagatgggggttatctactgataaagaagaattatgagaattgttgaaa agatggctaacaatctcggaagatttltgtttcttgggtttgttttttttttttttttt ta/tlctacagcttttatgaattcttaalgcttcaaaatgacttgggtctttctctttttttta catcagaatgggaataaagttaaaacccatagactcttcaaaactataggctagatagaaa tgtatgttgaactgttgaagctataatcagactatttaaatgttttgcatttttaactctaa aagattgtgctaatttattagagcagaacctgttggctctcctcagaagaagaactttccat tcaaatcacatggctttccccaataattttcaaaagataaattctgattttatgcaatggcatc	741
IGF1u1 8	IGF1	G	A	-	-	noncoding	GE683	TTTTATATTAC TGAGGCTTAA AGT	AGATATACCAT TTTTATTATGAC ACTCT	TTTTATTATCTAGAGGCTTAAAGTAAacattactctattttttggccaaaatgcactgatgt aaagtggaaaataaataaagagagctctaaatccctttcaagccaccattggcccccactcc aactcagcaaaagctactctgttaactccttaactgtatttggttggatatttactctgtac ccgtgctaacaacacactgcgagggagactctgaaacctaa[g/a]ctgtctacttactactcttt atctgtctgtgatacatgaaatgtctattcaaaatatacaaacctttcaaatatcacgcagc ttatattcagtttataaagggcccaaatccatgcagactcttttggtaaaaggttaactga actatgagaattgggattacatcatgtatttggctcattgtttttatccacttataggccaa gttgataaataaactacagacactgaatttaattccctgctactttgaaaccagaaaataat gcctggccattcgttacatctgtctgttgaagagcatatttttataaatttaattctgattg tattgaaattattatccaattcactatggcagaagaatccaatccctaatgactcttaaaat gtaactaatgaaatcattacttactatttactgttttaagacataatttgaataatgatggcta GAGTGCTATTAATAATGGTATATCT	741
IGF1u1 9	IGF1	C	G	-	-	noncoding	GE1191	TTTTATAGGAG TACATTTGAG AAC	ACAACTACAA AATACACCAT	TTTTATAGGAGTACATTTGATGAGACGcaagttagaggagtgaggaagaacaagaactacagatgt gggaagccctctctgagagtgagagtgacatgcccacccagagatccttgcctctgcacagctt acctgttaactttggaaacctctacaaaataaagttagatacaacttaaaagatggggtttc ccccatgaaaacacacaagtaaaacttccaacttggcttttggaggagatttgcaccttgcataa atggctctggagttgtgtgattgtgtattctttttatcaataatgttctatagaanaaaaaa aaataataataataataatacttctgctcctcctcaagagccacaatgctgggtgtgtgt ataatccagttgcactaaatt[c/g]ctctctgaatttgggtgctggagcattcattcagca acctgtcctaagggtttatgaatttttccattattggcaactcttctcacacactcgggtgt ttgtttacagtgctgataacttgttggcttatccccccccccttccataaactttatatt ttggcgaatttggccctcctcaagagcagcagtgctcagaagcagcacaacttttaaccac aagattccactctgtggcatttgcacaaataaagtggatgcatatttttttagacacaaact ttattttccacatcatgtttcaaaaaaagaatgaatgaatgttggcaacttttagggccaatca tttttaggcatagttttcaaacatagaagaatttctcaactcaaaagagttccttcaaatgta gttaatgcaaccttaattagtaacttctctcttttttttttttttttttttttttttttttt tttagcatacaactatcacaggatatacaacagtagtgaanaacctgttttttttagtataATGG TCTATTATTGTAGTTGT	993

FIG. 5KKKKK

FIG. 5L LLL

FIG. 5NNNNN

FIG. 500000

[illegible]

FIG. 5RRRR

[illegible]

FIG. 5SSSS

[illegible]

FIG. 5TTTTT

[illegible]

FIG. 5V VVV

133/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence
LIPCu9	LIPC	G	T	G	C	cds	GE323	CTCCGCGTAA CCTTTACC	CGGCCCATGAC TTTCACTTC	CTCCGCGTAACTTACCTGCTTCCCATAGGCTGGAGTCCGCGGAGCTTTGTTGAGG 295 gaagtcctcccaagcaatcgctcttccagatgattgctcaatgttggtgacattcatctatctt accgagagcacatggcgctgagcgtggatcaaacagccatagacacatgacatgactctatcc caacggggctctctccagcttg/tgctccacacttctagagctctacagacatattggccag cagggcttcattgggloagatgaaatgactgagggccg
MAOU1	MAOA	G	T	R	R	cds	GE1028	AGTGCATAATAC GTAATTAAATGC	AGCAGCCTTACC CTTCTTC	AGTGCATAATACGTAATTAAATGCGatccctccgaccttgactccaagatctcacttcagaccagag cttccagagagagcaaacacagtttaattcagcg/g/tctccaatggagagctgtcatttaagtgc tgatgattacaagaggcccttctggaaGAGAAGGTAGGCTGCT
MAOU2	MAOA	A	C	R	R	cds	GE984	GGGAAACAT ATTCATTT	ATTTCCTTCT TCCCAT	GGGAAACATATTCATTTGCGggggcccttccacagctatggatcccatctgcataattggatt 116 acaataactctgg/a/c/ggacaatagataaacATGCGAAGGTAAAT
MAOU1	MAOB	T	G	V	V	cds	GE1010	AAACAGAAC AAAGGTTAA	TGATCTTACT TTACATGGTG A	AAACAGAACCAAAAGGTTAAATatgtggaccttgagagatctctatg/t/g/ggaccaaccag aatcgtatcttgagattagccagagagctaggattggagacctacaagtgaatgggttgagcg tctgattCCACCATGTAAGGTAAAGGTAAATCA
MAOU2	MAOB	C	T	T	T	cds	GE1010	AAACAGAAC AAAGGTTAA	TGATCTTACT TTACATGGTG A	AAACAGAACCAAAAGGTTAAATatgtggaccttgagagatctctatg/tggaccaaccagaaac gtatcttgagattagccagagagctaggattggagac/c/t/tacaagtgaatgggttgagcg tctgattCCACCATGTAAGGTAAAGGTAAATCA
MAOU3	MAOB	C	T	P	P	cds	GE1068	TGAACCTAGGA TGTCCTG	CCCAACTCAT ATCCAAATAC A	TGAACCTAGGATGTCCTGcacagcccatccacacacctttttggagagacattggcc/c/t/t 225 ccgtccagcctgctcagctgattgattggaccacacatcttccagcaacggctcttggcttc ctggcccaaaaaggggctactgtgagagctctaaagagaggggtgtctgtaatacacactctc tctctactGTATTTGGGATATGAGTTTGGG
MAOU4	MAOB	A	G	K	K	cds	GE964	TTTAGATTAT CCTGGCCAC	TGTACTTTCC TCCTTGTTA	TTTAGATTATTCCTGGCCCAaagccagaaa/a/g/ctggcacgtctTACCAAGAGGAAAGGT 68 ACA
MAOU5	MAOB	A	G	E	E	cds	GE985	ATCTGACGCC AGTGCA	TCTACTTTCA TATTAGTCA G	ATCTGACGCCAGTGCAATTatgaga/a/g/laagaaactgggtgagagagagactctctgggggct 116 gctacacaaacttatttcccccctgggactCCTGACTCAATGGAAGGTAGA
MPLd14	MPL	G	A	-	-	noncoding	GE485	CTGTACAGTCC AGCCCTCC	GGTAGGAAGT TCACAGAGAG TC	CTGTACAGTCCAGCCCTCTccacaggaactgctttaatccagc/g/a cctctccatctctc 397 tcccagcccaagccacagctcagatccctgaaagagtggaacccagccctcttgaatcct ccccaaagctctcagagagagactccttccctgtgttctccacagccagatggactaccgaa gattcagccttcttctggggacacatgcccctgtgtgtgcccacccatggctgagcaggg tccgtgtacacacacattcccaacacatctaccataagctattggcagagccttg aggacaggctctcactccagcttccctggagagagactaaactctcagagactCTCTGTGGAACCT CCTTACC
MPLd15	MPL	G	A	P	P	cds	GE929	CTGGTCCCTC CCCTTC	CTGGAGTGGGA CTCAGC	CTGGTCCCTCTCTCacataaagctgctggagagccagggccaaactcaccagctgttcc 272 ttagatgtctcttggctggagacacagacacagccctaaagtgttctcccaaacatttgagga ctccactgtcttctgggagtgaggaagggcagccagtgggacatacaccagctgtgtatgctc accctg/a cggtaggtgtgagctgtgcccactccccatgatactgtccctgacattagCTGA GTCCACTCCAG
MPLd16	MPL	G	A	V	M	cds	GE917	AGAACTGGGC ATGGGC	GAATTCGGGAG CTGGAC	AGAACTGGGCATGGGccaggcttgggtctctcagggcgtcccgatggctgtgtagggggac 330 ctctctatgcccaaggggagagcccggtcttgcctccctgagctccctcagagcagcccaactt tggaaaccgatacgtgctccagcttcccaagagggagagtgctctctctctctccgtgacac tctgggtgagaaat/g/a tgttctaaacagactcagcagctcagctctcttcttgggacag tgtaggtaaagagcactcctctgctacccctgccccctccacttctgtgccccagctCCAGCTCCCG GAATC

FIG. 5XXXXX

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence	
MPLu1	MPL	C	T	A	V	cds	GE929	CTGGTCCCTC CCCTTC	CTGGAGTGGGA CTCAGC	CTGGTCCCTCCTCCCTCCacataaaccatgctggagagaccaggagccacacaccagctgttcc ttagatgtctctctgtgtggcatcagacacagagccctggaagtgttctccgaacattggagg ctctactgttctggagtgaggagggcag[c/t]ggccagtgggacataccacagctgtgtat gctaccgggtgagtgctgctgtgcccactcccattgctgtctctgcaacttaGCTGA GTCCACCTCCAG	272
MPLu10	MPL	A	G	T	A	cds	GE491	AGAGGCTGAGC CATAGACTGT	TGGGGCAAGAT TGAAGTAG	AGAGGCTGAGCCTAGACTGTggtaactcagagttctgagtgccctgtcttgcctcaggcctgc ggctccccccagtatcatcaagccatgggtgggagccagcagggggaacttcagatcagctgg gaggagccagctccagaaatcagtgatttcttgaggtacgaactccgctatggcccagagatcc caagaactcc[a/g]ctggctccaggtctacacagtaattgccacagaaactgtctccctgct ctgagagggcctcactcagctctgctctggaccagttctcatgtgacagcccaaatgcccctg gcaagatggaccacaaagacagctctcccagtagagagtagtgagcttcttgcctccacctc ttatctctTACTTCAATCTTGCCCCA	417
MPLu11	MPL	A	T	D	V	cds	GE490	GGGTTGGAGGC TCTCTCAG	CAGGCTTCCCT AGAGATATCT TTTA	GGGTTGGAGGCTCTCTCAGctgacagcagacacacagattgtgaagctgggatttctctcccaagg cttcagctctgacagcagagaggtggaggtgctctcctcaggtccagctgggaactctctac tggctgagctggcagggagacctg[a/t]tgggattctctctgggtgctctggggtctctggt cctctctgagctggagctgctggagtagagtagcaggtgagtcacacaaaggaataggagagtg gaggagatTAAAGAAATCTCTAGGGAGGCTG	293
MPLu12	MPL	C	A	A	A	cds	GE482	CTCTGTGTGGA CAATGCT	CCAGGATCCC CTGCGTA	CTCTGTGTGGAATGCTCTgtgacagagaggttaagctctcctctgctgacatcctctgagt ggctctccccccacccaaacttgctctggaggagatctccagtgggcatcttgaaattggagtggc agcctccatctctctggcagc[c/a]caagagaccttattcaactccgatacacagagagagg ccatcagagctggaaggtatggtcaagcaacaaatgccacagacctcacTACGCGGGATCCC TGG	263
MPLu13	MPL	T	A	L	Q	cds	GE450	CTCCTTGCCAA TCCACTG	AAGATCCAGT ACCAGGCAG	CTCCTTGCCAAATCCACTGccatggctcagctgctctctctctctctcccccagagagactgaggc atgcccctggccctcacttcagacctgacccgggtcc[a/a]aggccagtaaccttagggacac tgagcctgagcccggtgagtggtgtctctctctctctgcccacacacccCTGCTGGTACT GGATCCTT	203
MPLu2	MPL	C	A	P	Q	cds	GE491	AGAGGCTGAGC CATAGACTGT	TGGGGCAAGAT TGAAGGTAG	AGAGGCTGAGCCTAGACTGTggtaactcagagttctgagtgccctgtcttgcctcaggcctgc cggtccccccagtatcatcaagccatgggtgggagccagcagggggaacttcagatcagctgg gaggagc[c/a]agctccagaaatcagtgattctctgaggtacgaactccgtatggcccagag atcccagaactccactggtccacggtcatcagctgattgccacagaaactgtctccctgct ctgagagggcctcactcagctctgctctggaccgtctccatgtgctgagcccaaatgcccctg gcaagatggaccacaaagcagacctcccacagtagagagatgctgaccttcttgcctccacctc ttatctctTACTTCAATCTTGCCCCA	417
MPLu3	MPL	T	A	L	H	cds	GE491	AGAGGCTGAGC CATAGACTGT	TGGGGCAAGAT TGAAGGTAG	AGAGGCTGAGCCTAGACTGTggtaactcagagttctgagtgccctgtcttgcctcaggcctgc cggtccccccagtatcatcaagccatgggtgggagccagcagggggaacttcagatcagctgg gaggagccagctccagaaatcagtgattctctgaggtacgaacttcagatcagctgg atcccagaactccactggtccacggtcatcagctgattgccacagaaactgtctccctgct ctgagagggcctcactcagctctgctctggaccgtctccatgtgctgagcccaaatgcccctg gcaagatggaccacaaagcagacctcccacagtagagagatgctgaccttcttgcctccacctc ttatctctTACTTCAATCTTGCCCCA	417

FIG. 5YYYY

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
MPLu4	MPL	C	G	L	V	cds	GE482	CTCTGGTGACA CANTGCTT	CCAGGATCCC CTGCGTA	CTCTGGTGACAATGCTCTGTGCACAGAGACTTAAGCTGCTCCCTGCTGACATCCCTGTAGT gggctcccccaaccccaacttgcactggggagagatctccagtgggacat(c/g)tggaattggag tggcagcaagctgctctgggcagccagaccctgttatcaactcgcacacacagagagag ccatcaggactggaggtatggtcaagcaacaaatggccacagacacctcactACCGAGGGATCCC TGG
MPLu5	MPL	G	A	E	K	cds	GE491	AGAGGCTGAGC CATAGACTGT	TGGGGCAGAT TGAAGTAG	AGAGGCTGAGCCTAGACTGTGGTactcagagttctgtatgtgcctgtcttgcctcaggcctgc cggctcccccaagatcatcatcaagctggtgggagccagcaggggaacttcagatcagctgg gaggcagctccagaaatcagtgattcttcctgaggtac(g/a)aactcctgtatggccccag atccccagaactccactgtgtccacaggttcatacagttgacagtgatgcacagaaacctgtgcctgtct ctgcagaggtctcactcagcctctgtctggaccagttccatgtgctcagcccaaatgcctg gcaagatggaccacaaagcagacctcccccaagtagagaagatgtgacctctctgtgcctcaccctc ttatctctACCTTCATCTTGGCCCCA
MPLu6	MPL	C	A	R	R	cds	GE490	GGGTGGAGGC TCTCTCAG	CAGGCTCCCT AGAGATATTCT TTTA	GGGTGGAGGCTCTCTCAGctgacagggcagacctagattgtgaagctgggattttctctcccaagg cttcagctctgacagcagaggggtggaagctgcctcatctcaggaactcagcctggaactctctac tggctgaggtgg(c/a)agcgaactgattggatctcctctgggtgctctctgggagatcctggt ccctctctgtgactgtggacactgctctggagatggcaggtgggtcacaagaggaataggggagatggg ggggagatTAAGATATCTCTAGGGAGCCCTG
MPLu7	MPL	T	C	F	S	cds	GE472	ACGTGGGGCTG TATCTGACA	CAGGCTCCCT CTTCTG	ACGTGGGGCTGTATCTGACAggaactgagggctggcctgggaggggattggggcccaagcttcc tgaaggagatgggctgaaggcagggcacacagtgccggagagaagatgccctctctgggcccctct/ clcatggtcact gggtgacagaggggtggagatcacctatgcccACGAGAGGGAGCCCTG
MPLu8	MPL	C	T	S	S	cds	GE472	ACGTGGGGCTG TATCTGACA	CAGGCTCCCT CTTCTG	ACGTGGGGCTGTATCTGACAggaactgagggctggcctgggaggggattggggcccaagcttcc tgaaggagatgggctgaaggcagggcacacagtgccggagagaagatgccctctctgggcccctctca tggcacctc(c/r)tcct gggtgacagaggggtggagatcacctatgcccACGAGAGGGAGCCCTG
MPLu9	MPL	G	A	G	G	cds	GE491	AGAGGCTGAGC CATAGACTGT	TGGGGCAGAT TGAAGTAG	ACAGGCTGAGCCTAGACTGTGGTactcagagttctgtatgtgcctgtcttgcctcaggcctgc cggctcccccaagatcatcatcaagcctgggtgg(g/a)ggccagccaggggaacttcagatcag ctggggagggcagctccagaaatcagtgatttctctgaggtacgaactcctgtatggccccagag atccccagaactccactgtgtccacaggttcatacagttgacagaaacctgtgcctgtctctctct ctgcagaggtctcaactcagcctctgtctggaccagttctctatgtgtcagcccaaatgcctg gcaagatggaccacaaagcagacctcccccaagtagagaagatgtgacctctctgtgcctcaccctc ttatctctACCTTCATCTTGGCCCCA

FIG. 5ZZZZZ

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence	
NGFBu2	NGFB	A	G	-	-	noncoding	GE1186	TTTACAGAGGAGGCTGACGCTTTG	GACACAGCGGTAACCGG	TTTACAGAGGAGGCTGACGCTTTGCTACACATCTACAAGTATCATAGGAGCTCCGCGAGGCCAGTGGAGGCCCTCCAGGAGCAGCACTAATCCACAATCTAGTACCAAGTTGGGGATATTATTGTGGGTAACTGCAGTGCAGTATGGAGTCCCTCTGGGGCAGTTAGAGCCATACCATTTGATCTATAGTCCATAAGACAACAATAAAGAGAAAGACATGCTTAAGAGTGAAGAGAAAGAGGAGGAGGAGAAAGAGAGAGGAGTGGATGGAGGACACTAGTCTTAAGAGGCTCAACTTGGAACTCTATCTCTGTTCTGTTTCTTGTAGTCTTCTTGTAGTCTTCTTGTGAATCAGTTCCTTATATGTGAATAAATATGATAAATCCTAATGAACTCAGACTCATGAGATAGAGTGAAGTGAACACATTTAAACATCACACAAGAGGAACTATATGTGTCCACATTTATATGTGGGGTAGGCTCTGAAGGTCCTGGACTAAGTGGTCCAGAGCAAAAGTTTGTCCAAACATGACGCTTGTGAATCTACAAGGGGCTTGGTCCAGAGCATGAGGCTGACCCGTCAGCTGCTCGAAGGGGTCACAGTCTGTAGGCTTCAAGACATGCTCCAGCAGATCTCCCGTGCCTCCGAGAGATTCAAACTGTGAGCAGGACGACCATCACATCAGGACCAAGTCCAGGAGGAGGTGTTAACTCTCCCACTCCCTCTGTGACACATGACACTTACCCTCCCTCAGCGCTTAAAGTTCCAGAGAACTCAAGGACTCTGTAGTGTGTCCTCAAGTCTATCGAATCTGAGCAAAATTCAGGGGCTCTGCTCTCTCTGGAGAGCTGGGTGACCACACATCCATCTGCTGAGTCCAGCCCGGTTACGCTGTGTC	1002
NGFBu3	NGFB	G	T	-	-	noncoding	GE1186	TTTACAGAGGAGGCTGACGCTTTG	GACACAGCGGTAACCGG	TTTACAGAGGAGGCTGACGCTTTGCTACACATCTACAAGTATCATAGGAGCTCCGCGAGGCCAGTGGAGGCCCTCCAGGAGCAGCACTAATCCACAATCTAGTACCAAGTTGGGGATATTATTGTGGGTAACTGCAGTGCAGTATGGAGTCCCTCTGGGGCAGTTAGAGCCATACCATTTGATCTATAGTCCATAAGACAACAATAAAGAGAAAGACATGCTTAAGAGTGAAGAGAAAGAGGAGGAGGAGAAAGAGAGGAGTGGATGGAGGACACTAGTCTTAAGAGGCTCAACTTGGAACTCTATCTCTGTTCTGTTTCTTGTAGTCTTCTTGTAGTCTTCTTGTGAATCAGTTCCTTATATGTGAATAAATATGATAAATCCTAATGAACTCAGACTCATGAGATAGAGTGAAGTGAACACATTTAAACATCACACAAGAGGAACTATATGTGTCCACATTTATATGTGGGGTAGGCTCTGAAGGTCCTGGACTAAGTGGTCCAGAGCAAAAGTTTGTCCAAACATGACGCTTGTGAATCTACAAGGGGCTTGGTCCAGAGCATGAGGCTGACCCGTCAGCTGCTCGAAGGGGTCACAGTCTGTAGGCTTCAAGACATGCTCCAGCAGATCTCCCGTGCCTCCGAGAGATTCAAACTGTGAGCAGGACGACCATCACATCAGGACCAAGTCCAGGAGGAGGTGTTAACTCTCCCACTCCCTCTGTGACACATGACACTTACCCTCCCTCAGCGCTTAAAGTTCCAGAGAACTCAAGGACTCTGTAGTGTGTCCTCAAGTCTATCGAATCTGAGCAAAATTCAGGGGCTCTGCTCTCTCTGGAGAGCTGGGTGACCACACATCCATCTGCTGAGTCCAGCCCGGTTACGCTGTGTC	1002

FIG. 5B BBBBB

139/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/ noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence	
NGF-Bu6	NGF-B	T	C	-	-	noncoding	GE1186	TTTACAGAGGA GCTGACGTTTG	GACACAGCG TAACCCG	TTTACAGAGGAGCTGACGTTTCTACACATCTACAAGTATGCTAGGAGCTCCGCGGAGGCCAGT GAGAGCCCTCCAGGAGCAGCACTAATCCACAATAC[CTC]TGACCAAGTTGGGGATTAATTT GTGGGTAACTGCAAGTCAGTATGAGTCTCTCTGGGACAGTTAGAGCCATCACCATTGATCTCA CAGTCACATAGAACACACATAAAGAAAGACATCTCAGAGTGAAGAGAAAGAGGAGGAGGA GAAAGAGAGGGTGGATGGAGGACACTAGTAAGGGTCAACTTGGATTCATTC TGGTCAGTTTCTATTTGACTTCAGTCTTGTAGTCTTGTAGTCTTGTGATCAGTCTCC TATATGTGAATAAATATGATAAATCTTAATGAACACAGACTCAGAGTATGAGAGTGA ACACATTTAAACATCACAAAGAGGAATATTATGTGTCCACATTTATATGTGGTA GGTCTGAGAGGTGCTGGTAAAGATGGTCCAGAGCCCAAGGTTTTCACAAACATGAGC TTGTGAATCATAAAGGCTCCAGTCCAGTCCAGATCTTAGAGTGAACCTGACCTGCTCGA AAGGGTACCAGTTCTGGGCTCAAGCATGCCAGCAGATCTCCCGTGCCTCCCGA GAGTCAAACTGTAGCAGGAGGCACTCACATCAAGGCAAGTCCAGGAGGAGGT AACTCTCCCAACCTCTCTGTACACATGACACTTACCCTCCCTCAGCGCTTAA GCTCAGAGAACTCAAGGACTCTGTAAGTGTCTCCAGCTCATCGAAGTCTGCTGCAAA ATTGAGGGCTGTGACTCTCTGAGAGCTCGAGTGGGTGACCACATCCATCTGCTG AGTCAGCTCCGGTTACGCTCTGTCTC	1002
NGF-Bu7	NGF-B	C	T	-	-	noncoding	GE1186	TTTACAGAGGA GCTGACGTTTG	GACACAGCG TAACCCG	TTTACAGAGGAGCTGACGTTTCTACACATCTACAAGTATGCTAGGAGCTCCGCGGAGGCCAGT GAGAGCCCTCCAGGAGCAGCACTAATCCACAATAC[CTC]TGACCAAGTTGGGGATTAATTTGTGG GGTAACTGCAAGTCAGTATGAGTCTCTCTGGGACAGTTAGAGCCATCACCATTGATCTATAGT CACAATAGAACACACATAAAGAAAGACATCTTAAGAGTGAAGAGAAAGAGGAGGAGGA AGAAGAGAGGTGGATGGAGGACACTAGTATGAGGATCACTTGGATCTTCTCTGTGT TCAGTTTCTATTTGACTTCAGTCTTGTAGTCTTGTAGTCTTCTGTAATCAGTCTCTTAT ATGTGAATAAATATGATAAATCTTAATGAACCTCAGACTCATGAGAGTGAAGTGAAC ATTTAAACATCACAAAGAGGAATATTATGTGTCCACATTTATATATGTGGTGTAGCT CTGAGAGGTGCTGGACTAAGTGGTCCAGAGCCCAAGGTTTTCACAAACATGAGGCTTGG TGATCATAAAGGCTCCAGTCCAGCATCTCAGAGTCCAGGCTGAGTCTGCTGAGAGG GGTACAGTTCTGAGGCTCAAGACATCTCCCAAGCAGATCTCC[CT]TGTGCTCCCGA GGATCAAACTGTGAGGAGGAGGACCATCATCAAGTCAAGTCCAGGAGGAGGTGT AACTCTCCCAACCTCTCTGTACACATGACACTTACCCTCCCTCAGCGCTTAA GCTCAGAGAACTCAAGGACTCTGTAAGTGTCTCCAGCTCATCGAAGTCTGCTGCAAA ATTGAGGGCTGTGACTCTCTGAGAGCTCGAGTGGGTGACCACATCCATCTGCTG AGTCAGCTCCGGTTACGCTCTGTCTC	1002

FIG. 5DDDDDD

[illegible]

[illegible]

[illegible]

FIG. 5000000

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/ noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence	
PCI05	PCI	C	T	A	V	cds	GE413	GGACATCTCTG GAAATCAGC	TGAGGGAATTG GGTATTCCTTTA GAT	GGACATCTCTGGAAGTCAGCacctggacagctccaccctctctctgagagacacctcttccct ttcagacaagaagaccacatcgactctctctctctctgtgctgtggtgcttcttcagccctc agggggctcccttcacgcacaccaccccgagatgaagagagatcgagagacctccatga ggtcacacgggtggcccccagcagcagaaggacttacctttgacctcaaggcttcgactg cttcgctgccccagccagaacatctctctccctctggagcatctccaaggcttgccatg ctctccctggggactgggttcagacaaagtcgagatcctggagggcttgggctccaaactcca gaaaagctcagaagaagagctgcacagagcttcacgagctcttcaggaaactcaacagccca gaatggcttcagctgagctcgccatgctcttttcaccgactgggtggagactcgaggac accttcgaatgccatgaacgtgtacctcgagacacttccccaccaacttcaggactc tgcagggcccatgaagcatcaatgattatgtggcaagcaacgaaggccaagtgtggact tgtttaagaacctcgatagcaatcggtcgatcatgtgaattacatctcttttaaaagtgaag gcccttggggcccaaacctgcactcttttggctttctgctgcttttATCTTAAGAAATACCCAAAT TCCCTCA	787
PCI06	PCI	A	G	K	E	cds	GE413	GGACATCTCTG GAAATCAGC	TGAGGGAATTG GGTATTCCTTTA GAT	GGACATCTCTGGAAGTCAGCacctggacagctccaccctctctgagagacacttttccct ttcagacaagaagaccacatcgactctctctctctgtgctgtggtgcttcttcagccctc agggggctcccttcacgcacaccaccccgagatgaagagagatcgagagacctccatga gggtccacgggtggcccccagcagcagaaggacttacctttgacctcaaggcttcgacttc cgctgccccagccagaacatctctctccctgtggcatctccatgacctggccatgctctc ccctggggctgggttcagcacaagaatcgactctggagggcttgggctccactccagaaa acgtcagag (a/g) aggagctcagagagcttcagagctcttcaggaaactcaacagccca gagatggcttcagctgagctcgccatgctcttcaccgacttgggtgagactcgaggac accttcgaatgccatgaacgtgtacctcgagacacttccccaccaacttcaggactc tgcagggcccatgaagcatcaatgattatgtggcaagcaacgaaggccaagtgtggact tgtttaagaacctcgatagcaatcggtcgatcatgtgaattacatctcttttaaaagtgaag gcccttggggcccaaacctgcactcttttggctttctgctgcttttATCTTAAGAAATACCCAAAT TCCCTCA	787
PCI07	PCI	T	A	F	I	cds	GE413	GGACATCTCTG GAAATCAGC	TGAGGGAATTG GGTATTCCTTTA GAT	GGACATCTCTGGAAGTCAGCacctggacagctccaccctctctgagagacacttttccct ttcagacaagaagaccacatcgactctctctctctgtgctgtggtgcttcttcagccctc agggggctcccttcacgcacaccaccccgagatgaagagagatcgagagacctccatga gggtccacgggtggcccccagcagcagaaggacttacctttgacctcaaggcttcgacttc cgctgccccagccagaacatctctctccctgtggcatctccatgacctggccatgctctc ccctggggctgggttcagcacaagaatcgactctggagggctgggctccactccagaaa agctcagaagaagagctgcacagagcttccagcactcttcaggaaactcaacagcccaaga tggcttcagctgagctggcgaatggccttttcaacgacctgtgtgagactcagccggacact tggtaagtggcatgaagagctgacctggcagacacttccccaccaacttcaggagactctgca ggggccatgaagcagatcaatgattatgtggcaagcaacgaaggccaagtgtggacttggct taagaacctcgatagcaatcggtcgatcatgtgaattacatctcttttaaaagtgaag gcccttggggcccaaacctgcactcttttggctttctgctgcttttATCTTAAGAAATACCCAAAT TCCCTCA	787

FIG. 5TTTTT

[illegible]

FIG. 5 WWW

[illegible]

FIG. 5XXXXXX

[illegible]

FIG. 5YYYYYY

[illegible]

FIG. 5A

165/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence
SELP1 7	SELP	C	T	D	D	cds	GE454	TCACCTAGCT ATTTCGTGG	ACAGTACTTG ACAGACTGG	TCCACTAGCTATTTCGTGTGAGCGCTGGAGTCCCTGCCAGGAAGCATGGATTGCTCCCA tccttgagagcgtttcagctatga/c/tlaccactgtagcttcgctgctgctgaaggtttcagtc tgagagagcgcgataatagttcgtgtgtataacttggagacagtggagacagcagcccccagctctgt CAAGTACTGT
SELP1 8	SELP	A	T	S	C	cds	GE446	CTCTGCACTG AGAGAGTG	CACGTGTGGAG CTTATTG	CTCTGCACTGAGAGAGTGtggagaaacttgagcttgcctcccaacagctgctcatgaactgcagccac cctctgggaaacttcttcttaactgcagtcgagcttccactgcactgcagctgacgggtaccagtaaa tgggccc/a/t/gcagctgggaatgcttgggttctggatctggacaaATAAACCCTCCACACTG
SELP1 9	SELP	G	A	S	N	cds	GE466	TCTCTCCAGT GTSCAGT	AAATCTTACC CTCACAG	TCTCTCCAGTGTGCACTGtcagcacttggagagccccc[a/g/a]tgaagaaacacttggactgtgt tcctcgcctcactgcgttttgcctatggctccagctgcgaatttgagtgccagcccgctacagag tgaggggcttggacatgctccgctgcatgactctggacactgctgtgcaccccttgccaaCCTGT GAGGTAGATTTT
SELP2	SELP	T	C	N	N	cds	GE462	TACTCTAGCA TCAAGTGCC	ATATTATTACC TTTGCAGGTG	TACTCTAGCCATCAAGTCCcagaacttcttgcgccagagcagcgcctggattgttctgaca ctcgtggagaattcaatgttggtctccactgtcttctcttctgttaacaa[t/c]ggctttaagct ggaggggcccaataatgtggaatgcacaactcttggagatggttcagctactccacCAACTGCA AAGTAAATAAT
SELP2 0	SELP	C	G	S	*	cds	GE483	AACCAGAAAGA AGTGGCAG	AAGCCCTTACC TGTGTAA	AACCAGAAAGAAGTGGCAGcatggacttatcatcacagcacaagacatactcatggaatatttc ccgtaaataactgccagaacccgtacacagactttagtgccatccagaataaaatgaattgtt acctcaataaggctcctaccctactacagctcctactctggtatgggatccgaaagaaaca[t/a laagacatggacatgggtgggaaacccaaaggtctccacacagaggtgagacactgggctgata atgaacctaaacaaaggaacacagaggtcgtggagatacatcaagagtcctgcagcc cctggcaagtgggaatgatgagcactgttgaagaaagcagcattgtgtTACACAGGTAGGGC CTT
SELP2 1	SELP	T	A	N	K	cds	GE483	AACCAGAAAGA AGTGGCAG	AAGCCCTTACC TGTGTAA	AACCAGAAAGAAGTGGCAGcatggacttatcatcacagcacaagacatactcatggaatatttc ccgtaaataactgccagaacccgtacacagactttagtgccatccagaataaaatgaattgtt acctcaataaggctcctaccctactacagctcctactctggtatgggatccgaaagaaaca[t/a laagacatggacatgggtgggaaacccaaaggtctccacacagaggtgagacactgggctgata atgaacctaaacaaaggaacacagaggtcgtggagatacatcaagagtcctgcagcc cctggcaagtgggaatgatgagcactgttgaagaaagcagcattgtgtTACACAGGTAGGGC CTT
SELP2 2	SELP	T	A	C	S	cds	GE451	TCACAACAGGC ATAGCAT	CCTCTGCATGC TGGAGTT	TCACAACAGGCATAGCATcacttctcactccaggttgcgaatgtccagccctcaccactcctggg cagggaaacctgactgttaggcatactccgggaaccttggtttttaataccactgttactttgg ctgcaacgctgattccactcatagggagacagcactctcagc[t/a]gcagaccttcaggacaa tggacagcagtaACTCCAGCATGCGAGG
SELP3	SELP	T	A	N	K	cds	GE462	TACTCTAGCA TCAAGTGCC	ATATTATTACC TTTGCAGGTG	TACTCTAGCCATCAAGTCCcagaacttcttgcgccagagcagcgcctggattgttctgaca ctcgtggagaattcaatgttggtctccactgtcttctcttctgttaacaaatggctttaagctggag gggccc[a/t/a]aatgtggaatgcacaactcttggagatggttcagctactccacCAACTGCA AAGTAAATAAT
SELP4	SELP	T	G	S	A	cds	GE462	TACTCTAGCA TCAAGTGCC	ATATTATTACC TTTGCAGGTG	TACTCTAGCCATCAAGTCCcagaacttcttgcgccagagcagcgcctggattgttctgaca ctcgtggagaattcaatgttggtctccactgtcttctcttctgttaacaaatggctttaagctggag gggccc[a/t/a]aatgtggaatgcacaact[t/g]ctggagatggttcagctactccacCAACTGCA AAGTAAATAAT

FIG. 5DDDDDDDD

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/ noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
SELPu5	SELP	T	G	L	V	cds	GE451	TCACACAGGC ATAGCAT	CCCTCTGCATGC TGGAGTT	TCACACAGGCATGACATCACTTCTCTACTCCAGGTTG/gtgcaaatgtccagccctccacacactcc tgggcagggaaccatgactgtaggcatcatccgggaaccttgggttttaacacacactttgact ttggctgcacgctggatctacactcataggagacagcactctcagctgcacactccaggagaaa tggacagcagcgaACTCTCACATGCGNAGG
SELPu6	SELP	A	G	M	V	cds	GE466	TCTCTCCAGCT GTGCAGT	AAAATCTTACC CTCACAGG	TCTCTCCAGCTGTGCAGTGTGAGCACTGTGAAGCCCCAGTGAAGAACCATGGAGCTGTGTCAT cgcctcaactgcttttgcctatggctcagctgcacaaatttgatgtgcagccagctcagaggtgag ggctctggac/a/g/gtgcctgtgactctggacactgtgtcggacactgtgtcggcccttggccaaacCTGT GAGGTGAGATTTT
SELPu7	SELP	T	C	C	C	cds	GE462	TACTTAGCCA TCAATGCC	ATATTATTACC TTTGCAGGTG	TACTTAGCCATCAAGTGCAGCACTCTTCCCAGAGAGAGGAGAGCTGGATTGTTCTGACA ctctggagaattcaatgttgctctccacttg/c/cattctctctgtacaagtggtctcaagct tggggggcccaataatgtggaaatgcacactctctggagatggctcagctctccacCAACTGCA AAGTAAATAT
SELPu8	SELP	A	C	T	P	cds	GE452	CAGCTGTGAAA TGCTCAGA	AAAATTGTACC TTGCAGG	CAGCTGTGAATGCTCAGCACTacatgttaataagccaataggatgaactgctccacactctgg ggaaactcagttatggatcaatgtctctctctcattgtctagagggcagttacttaacagctc tgcaaaaaacagcatgccaaagagatggccacttgctca/a/c/ctacgtgtccaaCTTGCACAGGT ACAATTT
SELPu9	SELP	A	T	K	*	cds	GE483	AACCGAAGG AGTGCGAG	AAGCCCTTACC TGTGTAA	AACCGAAGGATGCGACGatggactctatcatcagccacaaaagacactcaatggaaatatttc ccgt/a/t/aatctgccgaatcgcctacacagatctgtggccacccagacactcaataaaat gatctactcaatgaatcctactactcagctctactactactgtgattggatccgaaagacaa taagacatggacatgggttgggacaaaaggctctccacaaagggtcgagacatggctgata atgaactcaacacaaaggaaacagagatcgctgtggagatatcatcaagagtcgctcagcc ctctggcagagtggaaatgatgagcactgcttgaagaaagacgcgactgtgtTACACAGGTAGGGC CTT
SHBGd3	SHBG	T	C	-	-	noncoding	GE539	TGACATGTGCC TACTCAGCTTT	AACACGGGAC TGGGTC	TGACATGTCCATCTACGACCTTgtttgtt/c/c/tctctctctctagatgagtgccacagccctcc ggctgtccactcagcaatggccagcagagagcctctgcgtgctcatgactcttggacctcaaca agatcaacaagatagggggttggcctagcccttGACCCAGTCCCTGGTT
SHBGd4	SHBG	C	T	P	L	cds	GE603	GCGAAGACAG ATCCGAG	CCAGAGTGTCT TTACCAAGTC	GCGAAGACAGATCCAGGGgacctctgtatttgcctccactctcctcagctggaagctcaagat ggagggggaactctgtcgtctggagtgagtgagagagagagagagagagagagagagagagag ggccctgaccacacaaacacatcccatggagatggccttggggggctgcctcttcccccct tccaaaccttgggttgc/c/199taactacacccaggggtgggaacccctagccaagACTTGGTAA AGACGTCTGG
SHBGd5	SHBG	G	A	D	N	cds	GE605	CACCTTAATGC TCTAATGCCA	TTGCTAGTSCA CAGTGAATATT TG	CACCTTAATGCTCTAATGCCAactttgcacactccctctccttagagaa/g/a/actcttccact cttttggatggagctttggccacaaaggtccagagctggatggagccagggccctcgaacaga agcatggatcctggactcagagctgccacagatgccacagatgccacagctgcctctccactta aagctccacctaaagacccctttgaaagtactgattattcatttaattcaCAAAATATTCACT GTGCATAGCAA
SHBGd6	SHBG	C	A	T	N	cds	GE605	CACCTTAATGC TCTAATGCCA	TTGCTAGTSCA CAGTGAATATT TG	CACCTTAATGCTCTAATGCCAactttgcacactccctctccttagagaaagactcttccacctttt ttgcctgaatggccttggccacaaaggtcagagctggatgtggacccagccctgaacagaagcc atgagatgtgc/a/1tcaagctgccccagagccagagctggagctgagctgcagctctccactta aagctccacctaaagacccctttgaaagtactgattattcatttaattcaCAAAATATTCACT GTGCATAGCAA

FIG. 5EEEEEE

[illegible]

[illegible]

FIG. 5

171/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence	
TBXAS1 a16	TBXAS 1	A	G	-	-	noncoding	GE332	CGAGATTGAAA TTTAAGGAAA GAC	CCAGAAACACA AGTGGTAACG A	CGAGATTGAAATTTAAGAAAGACAAAAGCTGTGAGATTGGGCTAACACGA [a/g]ctttctc cctttgtcagaccctccacacagatggccctggattcttcagctcgcagagagggccctgcctta tttgacatggtgattcgagacgctgagatgacccgagctttcaggtgtggtgagcc ctccctgcccagatcccccctacccctacccctgcccagctgaggtcagggccct ctccctCAGTTACACCTTGTTGTTCTGG	289
TBXAS1 a17	TBXAS 1	C	T	-	-	noncoding	GE332	CGAGATTGAAA TTTAAGGAAA GAC	CCAGAAACACA AGTGGTAACG A	CGAGATTGAAATTTAAGAAAGACAAAAGCTGTGAGATTGGGCTAACACGA [a/g]ctttctc tgcacagaccctccat [c/t]agatggccctgagttcttcagctcgcagagagggccctgcctta tttgacatggtgattcgagacgctgagatgacccgagctttcaggtgtggtgagcc ctccctgcccagatcccccctacccctacccctgcccagctgaggtcagggccct ctccctCAGTTACACCTTGTTGTTCTGG	289
TBXAS1 d12	TBXAS 1	C	G	Q	E	cds	GE357	TGCTGTTCCAA ATTGTTTACTG A	TTCAACACGCG AAAATCAAAAT	TGCTGTTCCAAATTTTACTGAAATagtttgaaatcttggaattttgcttaattcttctta tatagctgtgtttctctcaggggttttttggaaagc [c/g]aaatggagctcagaagctgta tggacctgtgtggtaagaagaacacacacgttcttatgtacagatatcttctattatgt acgatatTTTGGATTTCACGTGTTGAA	222
TBXAS1 d13	TBXAS 1	G	A	R	H	cds	GE326	TGGAACCTAT TCTTTTGCCCT T	TACAGCCATGA GCCACTGT	TGGAACCTATTTTTCCTTacttccagagagctcagtaattcttaggttctcctaagagcct aaagcatgagtgcaacttcatcttcagcttttgaatctgtcttctccctcaggtactccacat cagcatctccagactgagaagtaggctctcagacatcccaagccttctcttcatctggaac ttgacatttttcc [g/a]ccaggtgaagggtgtcttccatttggcttccatcaataatgctga ggccagggCACAGTGGCTCATGGCTGA	288
TBXAS1 d14	TBXAS 1	G	A	-	-	noncoding	GE355	CTTGGAGCATC CTTGTCTCA	GCTCTCACGCA GAGAACTGG	CTTGGAGCATCTTCTCTCAGatgcaggggtgtcagctgagacacagggctgcagggggagg gagcgggtgttctggccagccctgacacacagcagctgcaggttcaagctgcagggccgcagc agacggcccttccagctacctgccttggggccgcagcagcagctgcctcgggggtgcatcta ggctcttgaggtcaagttgacactgtctcacaagttccggttccaaagcctgccc tgagacacaggtgaagcccccctgctcagagggcag [g/a]tacaggggcagcgggtggagggcca ccccAGTCTCTGCGTGAGAGC	347
TBXAS1 u1	TBXAS 1	G	C	V	L	cds	GE355	CTTGGAGCATC CTTGTCTCA	GCTCTCACGCA GAGAACTGG	CTTGGAGCATCTTCTCTCAGatgcaggggtgtcagctgagacacagggctgcagggggagg gagcgggtgttctggccagccctgacacacagcagcagctgcaggttcaagctgcagggccgcagc agacggcccttccagctacctgccttggggccgcagcagcagctgcctcgggggtgcatcta ggctcttgaggtcaagttgacactgtctcacaagttccggttccaaagcct ggcctgagacacaggtgagggcccccctgctcagagggcaggtacaggggcagcgggtggagggcca ccccAGTCTCTGCGTGAGAGC	347
TBXAS1 u10	TBXAS 1	A	C	-	-	noncoding	GE249	ATGGACCTGTA TTGCCACCA	GAGAGTTTCCA TTTCTCAGTC TTA	ATGGACCTGTAATTTGCCACCAAGgttgggttccctcagctgagctgaacctctgcttgtt [a/c] cttcccaacagggcgtcgggtttgagtttggagttcagctgcgcagcaggttctgttttctacgtgac aaaagatgggaaggtcagaggtgcccctgctgctgtctcagttcagttgaaagctgaacgggt AAGACATGAGAAATGCAACTCTC	219
TBXAS1 u11	TBXAS 1	C	T	T	M	cds	GE355	CTTGGAGCATC CTTGTCTCA	GCTCTCACGCA GAGAACTGG	CTTGGAGCATCTTCTCTCAGatgcaggggtgtcagctgagacacagggctgcagggggagg gagcgggtgttctggccagccctgacacacagcagcagctgcaggttcaagctgcagggccgcagc agacggcccttccagctacctgccttggggccgcagcagcagctgcctcgggggtgcatcta tttagggctggttgaggttcaagttgacactgtctcacaagttccggttccaaagcct ggcctgagacacaggtgagggcccccctgctcagagggcaggtacaggggcagcgggtggagggcca ccccAGTCTCTGCGTGAGAGC	347

FIG. 5JJJJJJJJ

172/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
TBXAS1 u2	TBXAS 1	G	A	A	T	cds	GE274	CGACCTGGTGT TTCCCTCA	TGCTGCTCCA CTGGTAAT	CGACCTGGTGTTCCTCAGattcaacacggaggagcagctcaggaatgcgaggtgctggggcagcg catccccagcgc(g/a)ctgtgtagagatgctgggtgctgacatgacccctgagcac tgcccaagcccgagacattcaacccctgaaggctgagctgctcccttttaaaaagctctgaagg gatgtgagtggtggatagaaATTACAGTGGAGCGACGCA
TBXAS1 u3	TBXAS 1	C	T	A	A	cds	GE274	CGACCTGGTGT TTCCCTCA	TGCTGCTCCA CTGGTAAT	CGACCTGGTGTTCCTCAGattcaacacggaggagcagctcaggaatgcgaggtgctggggcagcg catccccagcgcgtgtgtagatggg(c/t)gtgggtgccctgacccatgacccctgagcac tgcccaagcccgagacattcaacccctgaaggctgagctgctcccttttaaaaagctctgaagg gatgtgagtggtggatagaaATTACAGTGGAGCGACGCA
TBXAS1 u4	TBXAS 1	C	G	Q	E	cds	GE274	CGACCTGGTGT TTCCCTCA	TGCTGCTCCA CTGGTAAT	CGACCTGGTGTTCCTCAGattcaacacggaggagcagct(c/g)aggactgcgaggtgctggggc agcgcacccccagcgcgtgtgtagatggccgtgggtgctgacccatgacccctgagcac tgcccaagcccgagacattcaacccctgaaggctgagctgctcccttttaaaaagctctgaagg gatgtgagtggtggatagaaATTACAGTGGAGCGACGCA
TBXAS1 u5	TBXAS 1	G	A	R	Q	cds	GE355	CTTGAGAGATC CTTGCTCA	GCTCTCAGCA GAGAACTGG	CTTGAGAGATCTTCCTCAGattgcaggggtggctcagctgagcagcagggctgcgaggggaggg gagcgggtgtctggcagcgcctgacacacgagcagctgcaggttcacggctgagggccggcagc agcacccgccttcaagtaacctgctcctgggcccgcacgagctgcctcgggtgcatcta ggcgtcttgaggtcaagtgcacactgctccagctgcacaaagtcc(g/a)gttccaaacct gcttgagcccgagtgaggccctctcagagcaggtacagggcgaggcggtgggggagggcca ccccagttctTCGGGTGAGAGC
TBXAS1 u6	TBXAS 1	T	G	V	G	cds	GE470	GGCCCTGGTTT ATTATCAC	CCAAAGTCGGC TCCATTC	GGCCCTGGTTTATTATCACcccttttcaatggccatttggttttctcttcaagtatctttc catcataatggtccactggccggattttggcccaataagaaacccgagacgaactgaatggcttt tttaacaaactcattaggaatg(t/g)gattgcttcggccagcagagctgccgaagaggttaa cgtattttaataggacacagccttgaattGAATGAGCGGACTTTGG
TBXAS1 u7	TBXAS 1	T	C	I	T	cds	GE912	GCCCATGTATC TTCCCTCCTTT	GGGGGATCAA CTTGTAAT	GCCCATGTATCTTCCTCCTTTgttctccaggaagcctcactctctgactgtaaggtcaaaatg tgcatttttctctcttcttcttagagggcagagacttctccaaatggctcctggatgccga catttgcgaagtcctatggcgtagaagatttgacatgtagagagctttctctctactg gtcaagccgaaccttccggcaacacccagccagcctatggccagcctttgacccgtggatg aga(t/c)tggtggcaggtctctctctcactgctggtgtatgaatcatcaccaacacact ttctttggccacctacactgcccacccctgactgctgagagagagctctctgagagaggtag acgtttttaaggagaaacagtgAGTACAGTTGGATCCCC
TBXAS1 u8	TBXAS 1	C	G	L	V	cds	GE912	GCCCATGTATC TTCCCTCCTTT	GGGGGATCAA CTTGTAAT	GCCCATGTATCTTCCTCCTTTgttctccaggaagcctcactctctgactgtaaggtcaaaatg tgcatttttctctcttcttcttagagggcagagacttctccaaatggctcctggatgccga catttgcgaagtcctatggcgtagaagatttgacatgtagagagctttctctctactg gtcaagccgaaccttccggcaacacccagccagcctatggccagcctttgacccgtggatg agattgtggccaggtctctctctcactgctggtgtatgaatcatcaccaacacacttct tttggcacctac(c/g)actggccacacacccctgactgctgagagagagctctctgagagaggtag acgtttttaaggagaaacagtgAGTACAGTTGGATCCCC
TBXAS1 u9	TBXAS 1	T	G	M	R	cds	GE274	CGACCTGGTGT TTCCCTCA	TGCTGCTCCA CTGGTAAT	CGACCTGGTGTTCCTCAGattcaacacggaggagcagctcaggaatgcgaggtgctggggcagcg catccccagcgcgtgtgtagatg(t/g)ggcgtgggtgccctgacccatgacccctgagcac tgcccaagcccgagacattcaacccctgaaggctgagctgctcccttttaaaaagctctgaagg gatgtgagtggtggatagaaATTACAGTGGAGCGACGCA

FIG. 5KKKKKKK

173/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/ noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
THBD1	THBD	T	G	C	G	cds	GE409	TCGGCTTACAG CTAATGTGC	GCACGCTAAGG TGCCTTTGGT	TCGGCTTACAGCTAATGTGCACCGCGCCGCGCCGCGGAGCGGTCCAGGGGCACTGGGCCAGGGAGGGCG CCGGGCGCTTGGGAGCTGAGCGTGGAGAACGGGGGTGGGAGCAGCGTGCATGCGATCCCTGG GGTCCCGCTGCCAGTGCAGCGCCGCGCCGCGCCGCGGAGCGCGCTCTGACCGCAT CCGCGCGAGTCCCTGCAACAGCACTCTGCGAGCACTCTGCGTCCCAACCGCCGACCGCGGGC TCTACTCGTGCATGTGGAGACCGGTACCGGTGCGCGCGGACCAACACCGGTGCGAGGACGT GGATGCTCATACTGGAGCCCGTCCGTCGCGAGCGTGTGCAACACAGGGTGGCTTCG AGTCCACTGTACCTAATCACTGCTGGTGGACGGGAGT/GTGTGGAGCCCTGGACCC GTCTTCAAGCCAACTCGAGTACCGTCCAGTCCAGCCCTGACCAACTAGTACTCTGTCT GCGCGAGGGCTTCGCGCCATTCCCAACGCGCGACAGGTCCAGATGTTTGCACCCAGCT GGCTGCCAGCGAGTGGACCCCAACCCAGGTAGTGTAGTCCCTGAAGGTACATCCT GGACGCGGTTTCACTGCAAGGACATCGACGAGTGCAGAACCGGCTCTGTCCCGGGTGT GGACAACTCCCGGTACCTTCGAGTGCATCTCGGGCCGCGCTGCGCCCTGCGCCCAT GGACCGAGTGTACTCCGCAAGTGGACGTTGGACAGCGTGGACAGCGCTCTGGAGCCCGCCAG CCGAGCGCGGTCCACTGACTCCTCCGCGGTGGGCTCGTGCATCGGGCTTGTCTCATAG GCATCTCCATCGCGAGCTGTGCTGGTGGTGGGCTTTGGGCTCCTCTGCCACCTGCGCAAG AAGCAGGGCGCGCGGAGGCAAGATGGAGTACAAGTGCAGCGCCCTT
THBD2	THBD	C	A	P	T	cds	GE409	TCGGCTTACAG CTAATGTGC	GCACGCTAAGG TGCCTTTGGT	TCGGCTTACAGCTAATGTGCACCGCGCCGCGCCGCGGAGCGGTCCAGGGGCACTGGGCCAGGGAGGGCG CCGGGCGCTTGGGAGCTGAGCGTGGAGAACGGGGGTGGGAGCAGCGTGCATGCGATCCCTGG GGTCCCGCTGCCAGTGCAGCGCCGCGCCGCGGAGCGCGCTCTGACCGCAT CCGCGCGAGTCCCTGCAACAGCACTCTGCGAGCACTCTGCGTCCCAACCGCCGACCGCGGGC TCTACTCGTGCATGTGGAGACCGGTACCGGTGCGCGCGGACCAACACCGGTGCGAGGACGT GGATGCTCATACTGGAGCCCGTCCGTCGCGAGCGTGTGCAACACAGGGTGGCTTCG AGTCCACTGTACCTAATCACTGCTGGTGGACGGGAGTGTGGAGCCGTGGACCCGTG TCAGAGCCAAGTGGAGTACAGTCAGT/C/a]CCCTGACCAACTAGTACTCTGTCT GCGCGAGGGCTTCGCGCCATTCCCAACGCGCGACAGGTCCAGATGTTTGCACCCAGCT GGCTGCCAGCGAGTGGACCCCAACCCAGGTAGTGTAGTCCCTGAAGGTACATCCT GGACGCGGTTTCACTGCAAGGACATCGACGAGTGCAGAACCGGCTCTGTCCCGGGTGT GGACAACTCCCGGTACCTTCGAGTGCATCTCGGGCCGCGCTGCGCCCTGCGCCCAT GGACCGAGTGTACTCCGCAAGTGGACGTTGGAGACGGCTCTGGAGCCCGCCAG CCGAGCGCGGTCCACTGACTCCTCCGCGGTGGGCTCGTGCATCGGGCTTGTCTCATAG GCATCTCCATCGCGAGCTGTGCTGGTGGTGGGCTTTGGGCTCCTCTGCCACCTGCGCAAG AAGCAGGGCGCGCGGAGGCAAGATGGAGTACAAGTGCAGCGCCCTT

FIG. 5L111111

174/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/ noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence	
THBDu3	THBD	C	T	P	P	cds	GB407	GACGACAGGA GAGCTCT	GTTGGGACG CAGAAGTG	GACGACAGGAGGCTGTGGCCATCGGCGTCTGTCCTCTGCTCCGCGCAGCGCCCTGTGCG AGTGCCTGCTTCCCCCGCGCTGCAAGCGGCGCTCGCTGGTAAATGCTTGGGTCTGGTCTC CTTGGCGTGGCTTCCCGCTGGGCTTCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG CCAGTGTGAGCACTGCTTGGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG AGATCTGCACTGCGGCG TCTTGTGCTGCACTGCGGCG ACCGGCTGCGGCG ACAACACGCTATAGCAGGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG TGCGTGTGCTGCTGCGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG CGAATGAAGCG TGAGCG GGGACTTCCAGCG TAATGTGCACTGCGGCG GACTGCACTGCGGCG CCAGTGCAGCG CGTGCACG CGCTGCACG CGCATCTCTGCTTCT ACCATCT CTTCTGAGCTTCT CCCTCTGCTGCGGCG CTGAAGGCT TACTGGCTTGGGCTTCT AACT ATCAGACAGGCT ACTGTTGACTTCT CTACTGGACGATAAGCTTCT CCCTGCTCTCTGAGCTTCT CACCCT CTC/GA/GTACAGCT	1014
THPOa6	THPO	G	A	-	-	noncoding	GB416	GGCATCTCTGTC TTTCTCTACTTA GAC	AGGAAATCTTG TCCAGTTGTCT C	GGCATCTCTGTCCTTCT ACCATCT CTTCTGAGCTTCT CCCTCTGCTGCGGCG CTGAAGGCT TACTGGCTTGGGCTTCT AACT ATCAGACAGGCT ACTGTTGACTTCT CTACTGGACGATAAGCTTCT CCCTGCTCTCTGAGCTTCT CACCCT CTC/GA/GTACAGCT	844
THPOa7	THPO	G	A	-	-	noncoding	GB416	GGCATCTCTGTC TTTCTCTACTTA GAC	AGGAAATCTTG TCCAGTTGTCT C	GGCATCTCTGTCCTTCT ACCATCT CTTCTGAGCTTCT CCCTCTGCTGCGGCG CTGAAGGCT TACTGGCTTGGGCTTCT AACT ATCAGACAGGCT ACTGTTGACTTCT CTACTGGACGATAAGCTTCT CCCTGCTCTCTGAGCTTCT CACCCT CTC/GA/GTACAGCT	844

FIG. 5MMMMMMMM

[illegible]

FIG. 5NNNNNNN

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence	
THP0u4	THPO	A	T	K	M	cds	GE416	GGCATCTCTGTC TTTCTTACTTAA GAC	AGGAAATCTTG TCCAGTGTCTC C	GGCATCTCTGTCCTTCTCTACTTACAGCAaggaggagcctgagatctcggccctgggtgttttgggcctcagg accatccttgccctcagcttctccacaggcggcaggaccacagctacaaaggatcccaatggccat cttctgagcttccacacctctcctcaggaa [a/t]gggtgggttctctctgtgtgtaggagg tccacctctggctcaggcggcccccaccacacagctgtccccagcagaacctctctagtccct cacactgaacagagctcccaaacaggacttctggattgttgagacaaactctcctgcctcagcca gaactactggcttgggtcttgaaatgggcagcagggtatccagcagcagatctcctgggtctctg aaccaacctccaggctccctggaccaaaccctggatctctcagacagatcacagaaactcttgaa tggactctggacttctctggacctcagcaggacctaggagcccgagcaatctctcag gaatcagacacagagctcctcgcacccaacctccagctgagatctctctccccacacat cctcctactggacagtatagctcttctctctccacacaccttgccccaccttgggtccagct ccacccctgctctcagacctctgtctccaaagcccccacctaccagcctcttctccaaacat cctacacccctcccaagatctgtctcaggaagggttaaggctctcagacacatcctccacacagca tctctcgtgtacagctctccttccctgcaggcggccctctgggagacacCTGGACAAAGATTTCCT	844
THu1	TH	G	A	S	S	cds	GE1125	GGTCCCCTGGGT CTCAGC	AGCCCCACCA CAGGTGA	GGTCCCCTGGGTCTCACaggttgaggagcattgggtgccccttgtccccacagctcccccggtcttc attgggcgcaggcagagcctcatcagagaccccagcaggagcgggagcggcggtggcagcagc ggcgtctcagctccctcggagcccgaggacccctcgaggctgtggcctttgagagaaggagg ggagggc [g/a]tgctaaacctgtctctccccaggggccacaaagccctcggcgctgtcccg agctgtgaagggtgttgaggtagctggtggcctctgtctccctgggggcaagtTCACCTGTGGGT GGGGCT	311
THu2	TH	G	A	V	M	cds	GE1125	GGTCCCCTGGGT CTCAGC	AGCCCCACCA CAGGTGA	GGTCCCCTGGGTCTCACaggttgaggagcattgggtgccccttgtccccacagctcccccggtcttc attgggcgcaggcagagcctcatcagagaccccagcaggagcgggagcggcggtggcagcagc ggcgtctcagctccctcggagcccgaggacccctcgaggctgtggcctttgagagaaggagg ggagggc [g/a]tgctaaacctgtctctccccaggggccacaaagccctcggcgctgtcccg agctgtgaagggtgttgaggtagctggtggcctctgtctccctgggggcaagtTCACCTGTGGGT GGGGCT	311
THu3	TH	T	G	F	C	cds	GE1125	GGTCCCCTGGGT CTCAGC	AGCCCCACCA CAGGTGA	GGTCCCCTGGGTCTCACaggttgaggagcattgggtgccccttgtccccacagctcccccggtcttc attgggcgcaggcagagcctcatcagagaccccagcaggagcgggagcggcggtggcagcagc ggcgtctcagctccctcggagcccgaggacccctcgaggctgtggcctttgagagaaggagg ggagggcgtgtctaaacctgtctctccccaggggccacaaagccctcggcgctgtcccg gtgaagatgt [t/g]tgaggtagctggtggcctctgtctccctgggggcaagtTCACCTGTGGGT GGGGCT	311
THu4	TH	G	A	K	K	cds	GE1020	CTGCCCGCAGG AAGGAG	CTGGGCACCC CTTCAG	CTGCCCGCAGGAAGgtgtcacacacgtgaa [g/a]ggcctctacgcacgcagcgcctcggtg ggagacactggaggcctttgtgtgagcgtctcagcgctaccgggaagagaacaatacccc agctggaggagcgtctccctctcCTGAAGGGTGTCCCCAG	170
THu5	TH	T	C	A	A	cds	GE1125	GGTCCCCTGGGT CTCAGC	AGCCCCACCA CAGGTGA	GGTCCCCTGGGTCTCACaggttgaggagcattgggtgccccttgtccccacagctcccccggtcttc attgggcgcaggcagagcctcatcagagaccccagcaggagcgggagcggcggtggcagcagc ggcgtctcagctccctcggagcccgaggacccctcgaggctgtggcctttgagagaaggagg ggagggcgtgtctaaacctgtctctccccaggggccacaaagccctcggcgctgtcccg gtgaagatgt [t/g]tgaggtagctggtggcctctgtctccctgggggcaagtTCACCTGTGGGT GGGGCT	331
THu6	TH	G	T	A	A	cds	GE972	TACGCGCAGGG ACTGCT	CGCTACCTGGG AGAACT	TACGCGCAGGGACTGCTgcccagagctgctgggggcagctgcccctggtgcccgcagccactctcg c [a/t]t [a/t]tCTGCTGCGAGGTACGC	89

FIG. 500000000

177/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence
VLDLR ₁₃	VLDLR	C	T	A	A	cds	GE926	GAAGAGCCTTGG CCTTCTTAAAG C	CAAGTGACAAT GACTTATGTCA AGA	GAAGAGCCTTGGCCTTCTTAAAGCAaaactaagtaaccacagacttccatctcttgaggaggaagagcc 281 aagctctgatctcactcaatcgaagagacatcaggaagattggcttagaggaagaataatattcc aacctagttgaacagctaaagaacacttggtctctcgatgctgacatgctgcccagaactattc tgggc[c/t]gatcgaagcaaaaggctatctcaggttaactttcagttcccttttgggtgTCTT GACATAAGTCATTCTCACCTTG
VLDLR ₁₄	VLDLR	A	T	-	-	noncoding	GE926	GAAGAGCCTTGG CCTTCTTAAAG C	CAAGTGACAAT GACTTATGTCA AGA	GAAGAGCCTTGGCCTTCTTAAAGCAaaactaagtaaccac[a/t]gacttccatcttcaggccaag 281 agccaagtctgatctcactcaatcgaagagacatcaggaagattggcttagaggaagaagaat atccaactagttgaacagctaaagaacacttggtctctcgatgctgacatgctgcccagaact attctggggcgatctcaagccaaggctatctcaggttaactttcagttcccttttgggtgTCTT GACATAAGTCATTCTCACCTTG
VLDLR ₁₅	VLDLR	G	T	-	-	noncoding	GE937	AGGTTTGGCT CCTTTACC	GGTAGCTCCAG ATGAACAAAA	AGGTTTGGCTCCTTACCTTgattgggttaaatttgaactgaatcacagatccttctaaactgatt 247 cctttatctctctgtagggatcaatgtaccacagcagatcagaggtcaggttcaggttcccccata gggacttctcgcgatgggcaattctctctctgtgaagta[g/t]attctctannngtctgggt caagaacttcttagatcaccagatgaagatTTTTTGTTCATCTGGAGCTACC
VLDLR ₁₆	VLDLR	G	A	E	K	cds	GE940	CCTGGGTTTA AATGTGAAGA TA	TATCCTTTCC ATCACCTGC	CCTGGGTTTAAATGTGAAGATATcaattgaaataaagttgtcaagtgantantacattttat 242 tcagataataaacgaatgcttggttaataatggttgatgttctcatctcgaagacacttagtt taggctac[g/a]agtgacttgagctgggtttggaactgagataggaagaaactctgagggg tgaactgaagaaagaaactgggacctGCAGGTGATGGGNAAGGATA
VLDLR ₁₇	VLDLR	C	T	N	N	cds	GE953	TTTTACAGCT TTGTTTACTGG T	TGAAGATAGTT GAGTGGGTGGT	TTTTACAGCTTTGTTTACTGCTgagactgggggtgaacacagctaaatagaaaaagcaggaaatga 176 attgattctgtagagctccactggtagcagcggaatctcagtggtcctaa[c/t]ggaaattacact tggatgtatgttctctctctcgcACCAACCACCTCAACTATCTTCA
VLDLR ₁	VLDLR	A	G	T	A	cds	GE920	GCTCTAATTTGT GTCAAACTCTT AAAT	GACCTACACAG ATACCATTTCA AAG	GCTCTAATTTGTGCTTAAATTTcttggacttattctgttccagtcctcaattgatga 324 caaggttggtagacatgttaaaatgatcgacaatgtctataatctcagcagcattgctgttgatt gggtgtacaagacacatctactggagctgagggctcttaagactattctcagtagctaccctagat ggg[a/g]ccaagaggagttctcttcaactctgacttgagagcgtcctccatagctgtggtg accacactgtctgggttctgtagtctgttttccatcacagCTTTGGAAATGGTATCTGTAGGTC
VLDLR ₁₀	VLDLR	A	T	S	C	cds	GE937	AGGTTTGGCT CCTTTACC	GGTAGCTCCAG ATGAACAAAA	AGGTTTGGCTCCTTACCTgattgggttaaatttgaactgaatcacagatccttctaaactgatt 247 cctttatctctctgtagggatcaatgtgacacagcaggtatcagaggtc[a/t]ggttccccc aaaaggacttctcgcgatgggcaattctctctctgtagtagatttctctannngtctgggt tcaagaacttcttagatcaccagatgaagatTTTTTGTTCATCTGGAGCTACC
VLDLR ₁₁	VLDLR	T	C	I	I	cds	GE937	AGGTTTGGCT CCTTTACC	GGTAGCTCCAG ATGAACAAAA	AGGTTTGGCTCCTTACCTgattgggttaaatttgaactgaatcacagatccttctaaactgatt 247 cctttatctctctgtagggatcaatgtgacacagcaggtatcagaggtcaggttctcccccata gggacttctcgcgatgggcaatt[c/t]ctctctctgtagtagatttctctannngtctgggt tcaagaacttcttagatcaccagatgaagatTTTTTGTTCATCTGGAGCTACC
VLDLR ₁₂	VLDLR	A	T	S	C	cds	GE945	TCCAATACTAG ACTTAGCTCAC TT	GACTTACTGCT GGGTACGTT	TCCAATACTAGCTTACTTACTTgattgggttaaatttgaactgaatcacagatccttctaaactgatt 228 cagtaggtggctacttgatgtggcggaattgggcaacacagaacatgaaaggcatgactttgac aatctctgacttgaacacacactgaagaggacctctccatagacattggtagacac[a/t]gtg cttctgttgggacacACCTACCCAGCAGTAGTAAGTC
VLDLR ₂	VLDLR	A	G	K	R	cds	GE926	GAAGAGCCTTGG CCTTCTTAAAG C	CAAGTGACAAT GACTTATGTCA AGA	GAAGAGCCTTGGCCTTCTTAAAGCAaaactaagtaaccacagacttccatctcttgaggga 281 agccaagtctgatctcactcaatcgaagagacatcaggaagattggcttagaggaagaagaat atccaactagttgaacagctaaagaacacttggtctctcgatgctgacatgctgcccagaact attctggggcgatctcaagccaaggctatctcaggttaactttcagttcccttttgggtgTCTT GACATAAGTCATTCTCACCTTG

FIG. 5PPPPPPP

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence	
VLDLRu 3	VLDLR	A	G	R	R	cds	GE920	GCTTAATGTTGT GTCAAACTCTT AAAT	GACCTACACAG ATACCATTCOA AAG	GCTTAATGTTGTGTCAAACCTTTAAATctcttgacactattctgtttccagtgccctcaattgatga caagggttggttag/a/g/catgtcttaaaatgcagcaaatgtctataaactctgcagccattgctgtt gatttggtgacagagaccactctactggactgcagagcttcttaagactcttaactgaacttcaatgaactaccct agatagaaccagagagagagttctctgttttaactgtgacttgagagagctgctccctccatcactgctggtg accctctgctgggtttggttagctgctgtttccatccacagcCTTGGAAATGATCTGTGTAGGTG	324
VLDLRu 4	VLDLR	A	G	Q	Q	cds	GE934	GAAATGGACTT GTGTTAATCCT G	TGCTGCTTCCC TTAAATAGTT AT	GAAATGGACTTGTGTTAACTCTGAGctacatctaatgtgggctctctgttttaggactgctc tactgagatagatggggaataatgaagcagctatagggcccaataaattaccctgacagctagcagctagc cactctagtcacacactgaaatgagccca/a/g/gacatcattgtctctatcatgacactgtcacag ccatcaggctaccgctggagagacagcccttaATAACTCTTTAAGGGAGACGCA	251
VLDLRu 5	VLDLR	A	C	E	A	cds	GE941	GCACCGGAATA CCCATTTTA	TTGCTTTGTTGT AGGTTCTACTT GTG	GCACCGGAATACCCATTTTAAATggtattttttctcctgactaggtataaaatgggtgctgaagaaga catgg/a/c/gaatggagagatggaatcactctgctgcgcacccacagattaatgactactct ccaaatataccctgtctctgctccagtggtgacaaatgtaggggaaatggccgagactgtcaaaag taaggcattttgtgtttccacCACAAAGTAGACCTTACACAGCAA	241
VLDLRu 6	VLDLR	C	T	L	F	cds	GE944	TGTACCTAGTA AGGTATAGGAG CAGC	CACCTTACTAT AAAGGTACAA AGCC	TGTACCTAGTAAGTATAGACACGCAgactaatctctgattctcctccagatattgatgaatg ccaaatcccggaatctgcagtcacaaatttgtatcacttaaaagggtttacaaatggaatga gtcgtggctatcaaatggat/c/t/ttctgactgctgctgcaaggcagtaggtaaatgaactgtg gactggtatGGCTGTGTACTCTTTATGATAGTGTG	230
VLDLRu 7	VLDLR	G	A	V	I	cds	GE236	ATTCTAGGGAG AAAAGCCAA	TTTACTTACCA CAGTTCTTTTC ATCA	ATTCTAGGGAGAAAAGCCAAAtgtaaacctcccaattccagtgacacaaatggctgctgtattac gctgtgtggaaatgtagtggggatgaagactgtg/a/ttgaaggcagTGAATAAGAACTGT GGTAAGTAA	140
VLDLRu 8	VLDLR	G	C	C	S	cds	GE911	CCAACTCTGAT GCATTTTCAG TG	CAGTTATACA GGGAAAGAAC TG	CCAACTCTGATGCAATTTTCAGggggcactctctctcttaataggcaataataacatgtagtcccg acgagttcaactgctccagtgccgctgactctccagaaactttgtatgaatggccagagatgac tgccagatggcagatgagctgagctgctgctcccaactgtgcccactgtgcccactgagttccagt/ g/c/cagcacctctctctgctatcccatcagctgggtatgagcagatgagcagactgctccgac caatctgatgagctccctggagcagtggtggcctcagcagctacacacacagtgctccagccag cgaatccagtgccgctctggcagtgatcccatagaagtggtggcagtgatgtaggggacccctgact gcaaggatggcagatgagggctcactgctgaagcagctctccagcagcagctggtcagctgttCAGTCTT TTCCTGTATCACTG	471
VLDLRu 9	VLDLR	T	A	I	N	cds	GE941	GCACCGGAATA CCCATTTTA	TTGCTTTGTTGT AGGTTCTACTT GTG	GCACCGGAATACCCATTTTAAAtggtattttttctcctgactaggtataaaatgggtggaagaaga catggagatggagagatgtgaatacctctatcctgcccagccacacaga/t/a/taatgactactct ccaaatataccctgtctctgctccagtggtgacaaatgtaggggaaatggccgagactgtcaaaag taaggcattttgtgtttccacCACAAAGTAGACCTTACACAGCAA	241

FIG. 5QQQQQQQ

RESULT 2

AAC71304

ID AAC71304 standard; DNA; 318 BP.

XX

AC AAC71304;

XX

DT 09-FEB-2001 (first entry)

XX

DE Single nucleotide polymorphism containing sequence #378.

XX

KW Single nucleotide polymorphism; SNP; human; genetic disease;
 KW disease susceptibility; cardiovascular system; endocrine system;
 KW neurological system; forensic testing; paternity testing; ds.

XX

OS Homo sapiens.

XX

PN WO200058519-A2. ✓

XX

PD 05-OCT-2000.

XX

PF 30-MAR-2000; 2000WO-US08440.

XX

PR 31-MAR-1999; 99US-0127248.

XX

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

PA (AFFY-) AFFYMETRIX INC.

XX

PI Altshuler D, Cargill M, Daley GQ, Ireland JS, Lander ES;

PI Lipshutz RJ, Patil N, Sklar P;

XX

DR WPI; 2000-611722/58.

XX

PT Nucleic acid selected from one of 106 genes comprising single
 PT nucleotide polymorphisms, allele-specific oligonucleotides to the genes
 PT are useful for phenotypic correlations, forensics, paternity testing,
 PT medicine and genetic analysis -

XX

PS Claim 1; Fig 5; 214pp; English.

XX

CC The present invention is concerned with a number of human single
 CC nucleotide polymorphisms (SNPs) which the inventors identified in human
 CC genes. These SNPs can be used in disease diagnosis and prediction of an
 CC individual's susceptibility to disease, in forensic and paternity testing
 CC and in genetic mapping. In particular, the SNPs of the invention can be
 CC used to diagnose susceptibility to diseases of the cardiovascular,
 CC endocrine and neurological systems, such as coronary artery disease,
 CC schizophrenia, cancer, autoimmune diseases, Alzheimer's and Parkinson's
 CC diseases.

CC Note: The degenerate codon within the sequence represents the position
 CC of an SNP, for example the letter S represents a polymorphism where the
 CC nucleotide may be C or G.

XX

SQ Sequence 318 BP; 66 A; 106 C; 90 G; 55 T; 1 other;

Query Match 99.8%; Score 200.6; DB 21; Length 318;

Best Local Similarity 99.5%; Pred. No. 1.7e-47;

Matches 200; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	TCCCCAGACAAGGATGACCAGCTGATCTGTGTGAACGAGAACGGCGGCTGTGAGCAGTAC	60
Db	94	TCCCCAGACAAGGATGACCAGCTGATCTGTGTGAACGAGAACGGCGGCTGTGAGCAGTAC	153
Qy	61	TGCAGTGACCACACGGGCACCAAGCGCTCCTGTCCGTGCCACGAGGGGTACTCTCTGCTG	120
Db	154	TGCAGTGACCACACGGGCACCAAGCGCTCCTGTCCGTGCCAYGAGGGGTACTCTCTGCTG	213
Qy	121	GCAGACGGGGTGTCTTGCACACCCACAGGTGACCAGGCTTCATGTCCCAGTCCCAGATGA	180
Db	214	GCAGACGGGGTGTCTTGCACACCCACAGGTGACCAGGCTTCATGTCCCAGTCCCAGATGA	273
Qy	181	CACCAGTCCCTGTCCCACTAG	201
Db	274	CACCAGTCCCTGTCCCACTAG	294

RESULT 1

AAC71295

ID AAC71295 standard; DNA; 266 BP.

XX

AC AAC71295;

XX

DT 09-FEB-2001 (first entry)

XX

DE Single nucleotide polymorphism containing sequence #375.

XX

KW Single nucleotide polymorphism; SNP; human; genetic disease;

KW disease susceptibility; cardiovascular system; endocrine system;

KW neurological system; forensic testing; paternity testing; ds.

XX

OS Homo sapiens.

XX

PN WO200058519-A2.

XX

PD 05-OCT-2000.

XX

PF 30-MAR-2000; 2000WO-US08440.

XX

PR 31-MAR-1999; 99US-0127248.

XX

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

PA (AFFY-) AFFYMETRIX INC.

XX

PI Altshuler D, Cargill M, Daley GQ, Ireland JS, Lander ES;

PI Lipshutz RJ, Patil N, Sklar P;

XX

DR WPI; 2000-611722/58.

XX

PT Nucleic acid selected from one of 106 genes comprising single

PT nucleotide polymorphisms, allele-specific oligonucleotides to the genes

PT are useful for phenotypic correlations, forensics, paternity testing,

PT medicine and genetic analysis -

XX

PS Claim 1; Fig 5; 214pp; English.

XX

CC The present invention is concerned with a number of human single

CC nucleotide polymorphisms (SNPs) which the inventors identified in human

CC genes. These SNPs can be used in disease diagnosis and prediction of an

CC individual's susceptibility to disease, in forensic and paternity testing

CC and in genetic mapping. In particular, the SNPs of the invention can be

CC used to diagnose susceptibility to diseases of the cardiovascular,

CC endocrine and neurological systems, such as coronary artery disease,

CC schizophrenia, cancer, autoimmune diseases, Alzheimer's and Parkinson's

CC diseases.

CC Note: The degenerate codon within the sequence represents the position

CC of an SNP, for example the letter S represents a polymorphism where the

CC nucleotide may be C or G.

XX

SQ Sequence 266 BP; 48 A; 84 C; 78 G; 55 T; 1 other;

Query Match 99.8%; Score 200.6; DB 21; Length 266;
Best Local Similarity 99.5%; Pred. No. 6.8e-48;

Matches 200; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 GCAGCACTGCAGAGATTTTCATCATGGTCTCCCAGGCCCTCAGGCTCCTCTGCCTTCTGCT 60
      |||
Db      59 GCAGCACTGCAGAGATTTTCATCATGGTCTCCCAGGCCCTCAGGCTCCTCTGCCTTCTGCT 118

Qy      61 TGGGCTTCAGGGCTGCCTGGCTGCAGGTGCGTCCGGGGAGGTTTTCTCCATAAACTTGGT 120
      |||
Db     119 TGGGCTTCAGGGCTGCCTGGCTGCAGGTGCGTCCRGGGAGGTTTTCTCCATAAACTTGGT 178

Qy     121 GGAAGGGCAGTGGGCAAATCCAGGAGCCAGCCCGGGCTTCCCAAACCCCGCCCTTGCTCC 180
      |||
Db     179 GGAAGGGCAGTGGGCAAATCCAGGAGCCAGCCCGGGCTTCCCAAACCCCGCCCTTGCTCC 238

Qy     181 GGACACCCCCATCCACCAGGA 201
      |||
Db     239 GGACACCCCCATCCACCAGGA 259
```